(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
29 April 2004 (29.04.2004)

PCT

(10) International Publication Number
WO 2004/035732 A2

(51) International Patent Classification[7]: **C12N**

(21) International Application Number:
PCT/US2003/026780

(22) International Filing Date: 28 August 2003 (28.08.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/406,579	29 August 2002 (29.08.2002)	US
60/406,642	29 August 2002 (29.08.2002)	US
60/406,588	29 August 2002 (29.08.2002)	US
60/406,646	29 August 2002 (29.08.2002)	US
60/406,653	29 August 2002 (29.08.2002)	US
60/406,608	29 August 2002 (29.08.2002)	US
60/406,585	29 August 2002 (29.08.2002)	US
60/406,616	29 August 2002 (29.08.2002)	US
60/406,655	29 August 2002 (29.08.2002)	US
60/406,640	29 August 2002 (29.08.2002)	US
60/406,576	29 August 2002 (29.08.2002)	US
60/406,666	29 August 2002 (29.08.2002)	US
60/406,611	29 August 2002 (29.08.2002)	US
60/406,612	29 August 2002 (29.08.2002)	US
60/410,947	17 September 2002 (17.09.2002)	US
60/410,958	17 September 2002 (17.09.2002)	US
60/411,035	17 September 2002 (17.09.2002)	US
60/410,961	17 September 2002 (17.09.2002)	US
60/411,045	17 September 2002 (17.09.2002)	US
60/411,041	17 September 2002 (17.09.2002)	US
60/410,959	17 September 2002 (17.09.2002)	US
60/411,023	17 September 2002 (17.09.2002)	US
60/411,055	17 September 2002 (17.09.2002)	US
60/411,048	17 September 2002 (17.09.2002)	US
60/411,073	17 September 2002 (17.09.2002)	US
60/411,019	17 September 2002 (17.09.2002)	US
60/411,024	17 September 2002 (17.09.2002)	US
60/411,046	17 September 2002 (17.09.2002)	US
60/411,082	17 September 2002 (17.09.2002)	US
60/411,022	17 September 2002 (17.09.2002)	US
60/410,962	17 September 2002 (17.09.2002)	US
60/410,953	17 September 2002 (17.09.2002)	US
60/410,957	17 September 2002 (17.09.2002)	US
60/411,037	17 September 2002 (17.09.2002)	US
60/411,032	17 September 2002 (17.09.2002)	US
60/410,946	17 September 2002 (17.09.2002)	US
60/410,960	17 September 2002 (17.09.2002)	US
60/411,111	17 September 2002 (17.09.2002)	US
60/411,052	17 September 2002 (17.09.2002)	US
60/410,949	17 September 2002 (17.09.2002)	US
60/410,948	17 September 2002 (17.09.2002)	US
60/411,101	17 September 2002 (17.09.2002)	US
60/463,732	18 April 2003 (18.04.2003)	US
60/463,700	18 April 2003 (18.04.2003)	US
60/463,708	18 April 2003 (18.04.2003)	US
60/463,716	18 April 2003 (18.04.2003)	US
60/467,199	2 May 2003 (02.05.2003)	US
60/467,201	2 May 2003 (02.05.2003)	US
60/467,230	2 May 2003 (02.05.2003)	US
60/467,203	2 May 2003 (02.05.2003)	US
60/471,306	19 May 2003 (19.05.2003)	US
60/471,336	19 May 2003 (19.05.2003)	US
60/472,430	22 May 2003 (22.05.2003)	US
60/472,420	22 May 2003 (22.05.2003)	US
60/476,609	9 June 2003 (09.06.2003)	US
60/476,641	9 June 2003 (09.06.2003)	US
60/485,224	8 July 2003 (08.07.2003)	US
60/485,218	8 July 2003 (08.07.2003)	US
60/485,223	8 July 2003 (08.07.2003)	US
60/485,325	8 July 2003 (08.07.2003)	US
60/486,480	14 July 2003 (14.07.2003)	US
60/486,446	14 July 2003 (14.07.2003)	US
60/486,891	15 July 2003 (15.07.2003)	US
60/486,960	15 July 2003 (15.07.2003)	US
60/493,573	8 August 2003 (08.08.2003)	US
60/493,370	8 August 2003 (08.08.2003)	US
60/493,341	8 August 2003 (08.08.2003)	US
60/493,577	8 August 2003 (08.08.2003)	US

(71) Applicant (for all designated States except US): FIVE PRIME THERAPEUTICS, INC. [US/US]; 951 Gateway Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): WILLIAMS, Lewis, T. [US/US]; 125 Chapel Drive, Mill Valley, CA 94941 (US). CHU, Keting [US/US]; 2017 Easton Drive, Burlingame, CA 94010 (US). LEE, Ernestine [US/US]; 226 Arlington Avenue, Kensington, CA 94707 (US). HES-TIR, Kevin [US/US]; 226 Arlington Avenue, Kensington, CA 94707 (US). BEAURANG, Pierre, Alvaro [BE/US]; 653 Via Casitas, Greenbrae, CA 94904 (US). BEHRENS, Dirk [DE/US]; 5 Delmar Court, Redwood City, CA 94063 (US). HALENBECK, Robert, Forgan [US/US]; 59

(54) Title: HUMAN POLYPEPTIDES ENCODED BY POLYNUCLEOTIDES AND METHODS OF THEIR USE

(57) Abstract: The invention provides novel polynucleotides, related polypeptides related nucleic acid and polypeptide compositions, and related modulators, such as antibodies and small molecule modulators. The invention also provides methods to make and use these polynucleotides, polypeptides, related compositions, and modulators. These methods include diagnostic, prophylactic and therapeutic applications. The compositions and methods of the invention are useful in treating proliferative disorders, e.g., cancers, and inflammatory, immune, bacterial, and viral disorders.

Golden Hinde Boulevard, San Rafael, CA 94903 (US).
HUANG, Min, Mei [US/US]; 15682 Anchorage Drive,
San Leandro, CA 94579 (US). **KOTHAKOTA, Srinivas**
[IN/US]; 932 7th Street, Apt. 10, Santa Monica, CA
90403 (US). **HAISHAN, Lin** [US/US]; 20942 Sherman
Drive, Castro Valley, CA 94552 (US). **LINNEMANN,
Thomas** [DE/US]; 1338 Baker St. #7, San Francisco, CA
94115 (US). **PIERCE, Kristen** [US/US]; 1336 Capuchino
Avenue, Burlingame, CA 94010 (US). **WANG, Yan**
[CN/US]; 1075 Park Place, #306, San Mateo, CA 94403
(US). **WONG, Justin, G., P.** [US/US]; 389 Somerset Road
#3, Oakland, CA 94611 (US). **WU, Ge** [US/US]; 954
Brentwood Drive, San Jose, CA 95129 (US). **ZHANG,
Hongbing** [CN/US]; 925 Cerrito Street, Albany, CA
94706 (US).

(74) **Agent: GARRETT, Arthur, S.;** Finnegan, Henderson,
Farabow, Garrett & Dunner, L.L.P., 1300 I Street, N.W.,
Washington, DC 20005-3315 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC,
SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,
SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished
upon receipt of that report*
— *with sequence listing part of description published sepa-
rately in electronic form and available upon request from
the International Bureau*

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

HUMAN POLYPEPTIDES ENCODED BY
POLYNUCLEOTIDES AND METHODS OF THEIR USE

PRIORITY CLAIM

[001]   This application is related to the following provisional applications filed in the United States Patent and Trademark Office, the disclosures of which are hereby incorporated by reference:

Application Number	Title	Filing Date
60/406,616	Polynucleotides Encoding Secreted Proteins and Secreted Proteins Encoded Thereby	August 29, 2002
60/406,579	Methods of Use for Polynucleotides Encoding Secreted Proteins and Secreted Proteins Encoded Thereby	August 29, 2002
60/406,655	Polynucleotides Encoding Single Transmembrane Proteins And Single Transmembrane Proteins Encoded Thereby	August 29, 2002
60/406,642	Methods of Use for Polynucleotides Encoding Single Transmembrane Proteins and Single Transmembrane Proteins Encoded Thereby	August 29, 2002
60/406,640	Polynucleotides Encoding Multiple Transmembrane Proteins And Multiple Transmembrane Proteins Encoded Thereby	August 29, 2002
60/406,588	Methods of Use for Polynucleotides Encoding Multiple Transmembrane Proteins and Multiple Transmembrane Proteins Encoded Thereby	August 29, 2002
60/406,576	Polynucleotides Encoding Kinases and Kinases Encoded Thereby	August 29, 2002
60/406,646	Methods of Use for Polynucleotides Encoding Kinases and Kinases Encoded Thereby	August 29, 2002
60/406,666	Polynucleotides Encoding Proteases and Proteases Encoded Thereby	August 29, 2002

60/406,653	Methods Of Use for Polynucleotides Encoding Proteases And Proteases Encoded Thereby	August 29, 2002
60/406,611	Polynucleotides Encoding Phosphatases and Phosphatases Encoded Thereby	August 29, 2002
60/406,608	Methods of Use for Polynucleotides Encoding Phosphatases and Phosphatases Encoded Thereby	August 29, 2002
60/406,612	Polynucleotides Encoding Polypeptides and Polypeptides Encoded Thereby	August 29, 2002
60/406,585	Methods of Use for Polynucleotides Encoding Polypeptides And Polypeptides Encoded Thereby	August 29, 2002
60/411,019	Polynucleotides Encoding Secreted Proteins and Secreted Proteins Encoded Thereby	September 17, 2002
60/411,024	Novel Polynucleotides Encoding Secreted Proteins and Novel Secreted Proteins Encoded Thereby	September 17, 2002
60/410,947	Methods of Use for Polynucleotides Encoding Secreted Proteins and Secreted Proteins Encoded Thereby	September 17, 2002
60/410,958	Methods of Use for Novel Polynucleotides Encoding Secreted Proteins and Novel Secreted Proteins Encoded Thereby	September 17, 2002
60/411,046	Polynucleotides Encoding Single Transmembrane Proteins and Single Transmembrane Proteins Encoded Thereby	September 17, 2002
60/411,082	Novel Polynucleotides Encoding Single Transmembrane Proteins and Novel Single Transmembrane Proteins Encoded Thereby	September 17, 2002
60/411,035	Methods of Use for Polynucleotides Encoding Single Transmembrane Proteins and Single Transmembrane Proteins Encoded Thereby	September 17, 2002
60/410,961	Methods of Use for Novel Polynucleotides Encoding Single Transmembrane Proteins and Novel Single Transmembrane Proteins Encoded Thereby	September 17, 2002

60/411,022	Polynucleotides Encoding Multiple Transmembrane Proteins and Multiple Transmembrane Proteins Encoded Thereby	September 17, 2002
60/410,962	Novel Polynucleotides Encoding Multiple Transmembrane Proteins and Novel Multiple Transmembrane Proteins Encoded Thereby	September 17, 2002
60/411,045	Methods of Use for Polynucleotides Encoding Multiple Transmembrane Proteins and Multiple Transmembrane Proteins Encoded Thereby	September 17, 2002
60/411,041	Methods of Use for Novel Polynucleotides Encoding Multiple Transmembrane Proteins and Novel Multiple Transmembrane Proteins Encoded Thereby	September 17, 2002
60/410,953	Polynucleotides Encoding Kinases And Kinases Encoded Thereby	September 17, 2002
60/410,957	Novel Polynucleotides Encoding Kinases and Novel Kinases Encoded Thereby	September 17, 2002
60/410,959	Methods of Use for Polynucleotides Encoding Kinases and Kinases Encoded Thereby	September 17, 2002
60/411,023	Methods of Use for Novel Polynucleotides Encoding Kinases and Novel Kinases Encoded Thereby	September 17, 2002
60/411,037	Polynucleotides Encoding Phosphatases and Phosphatases Encoded Thereby	September 17, 2002
60/410,951	Novel Polynucleotides Encoding Phosphatases and Novel Phosphatases Encoded Thereby	September 17, 2002
60/410,948	Methods of Use for Polynucleotides Encoding Phosphatases and Phosphatases Encoded Thereby	September 17, 2002
60/410,949	Methods of Use for Novel Polynucleotides Encoding Phosphatases and Novel Phosphatases Encoded Thereby	September 17, 2002
60/410,946	Polynucleotides Encoding Proteases and Proteases Encoded Thereby	September 17, 2002
60/410,960	Novel Polynucleotides Encoding Proteases and Novel Proteases Encoded Thereby	September 17, 2002

60/411,055	Methods of Use for Polynucleotides Encoding Proteases and Proteases Encoded Thereby	September 17, 2002
60/411,048	Methods of Use for Novel Polynucleotides Encoding Proteases and Novel Proteases Encoded Thereby	September 17, 2002
60/411,111	Polynucleotides Encoding Polypeptides and Polypeptides Encoded Thereby	September 17, 2002
60/411,052	Novel Polynucleotides Encoding Proteins and Novel Proteins Encoded Thereby	September 17, 2002
60/411,073	Methods of Use for Polynucleotides Encoding Polypeptides and Polypeptides Encoded Thereby	September 17, 2002
60/411,101	Methods of Use for Novel Polynucleotides Encoding Polypeptides and Novel Polypeptides Encoded Thereby	September 17, 2002

TECHNICAL FIELD

[002]   The present invention is related generally to novel polynucleotides and novel polypeptides encoded thereby, their compositions, antibodies directed thereto, and other agonists or antagonists thereto. The polynucleotides and polypeptides are useful in diagnostic, prophylactic, and therapeutic applications for a variety of diseases, disorders, syndromes, and conditions, as well as in discovering new diagnostics, prophylactics, and therapeutics for such diseases, disorders, syndromes, and conditions (hereinafter disorders). The present invention also relates to methods of modulating biological activities through the use of the novel polynucleotides and novel polypeptides of the invention and through the use of agonists and antagonists, such as antibodies, thereto.

[003]   This application further relates to the field of polypeptides that are associated with regulating cell growth and differentiation, that are over-expressed in cancer, and/or that can be associated with proliferation or inhibition of cancer growth, including hematopoietic cancers such as leukemias, lymphomas, and solid cancers such as lung cancer, for example, adenocarcinomas and/or squamous cell carcinomas. These polypeptides may also be associated with other conditions, such as inflammatory, immune, and metabolic disorders, as well as microbial infections, including viral, bacterial, fungal, and parasitic disorders.

4

[004]   This application further relates to modulators of biological activity that can specifically bind to these polynucleotides or polypeptides, or otherwise specifically modulate their activity. For example, they can directly or indirectly induce antibody-dependent cellular cytotoxicity (ADCC), complement-dependent cytotoxicity (CDC), endocytosis, apoptosis, or recruitment of other cells to effect cell activation, cell inactivation, cell growth or differentiation or inhibition thereof, and cell killing.

[005]   The sequences of the invention encompass a variety of different types of nucleic acids and polypeptides with different structures and functions. They can encode or comprise polypeptides belonging to different protein families ("Pfam"). The "Pfam" system is an organization of protein sequence classification and analysis, based on conserved protein domains; it can be publicly accessed in a number of ways, for example, at http://pfam.wustl.edu. Protein domains are portions of proteins that have a tertiary structure and sometimes have enzymatic or binding activities; multiple domains can be connected by flexible polypeptide regions within a protein. Pfam domains can comprise the N-terminus or the C-terminus of a protein, or can be situated at any point in between. The Pfam system identifies protein families based on these domains and provides an annotated, searchable database that classifies proteins into families (Bateman et al., 2002).

[006]   Sequences of the invention can encode or be comprised of more than one Pfam. Sequences encompassed by the invention include, but are not limited to, the polypeptide and polynucleotide sequences of the molecules shown in the Sequence Listing and corresponding molecular sequences found at all developmental stages of an organism. Sequences of the invention can comprise genes or gene segments designated by the Sequence Listing, and their gene products, i.e., RNA and polypeptides. They also include variants of those presented in the Sequence Listing that are present in the normal physiological state, e.g., variant alleles such as SNPs, splice variants, as well as variants that are affected in pathological states, such as disease-related mutations or sequences with alterations that lead to pathology, and variants with conservative amino acid changes. Sequences of the invention are categorized below; any given sequence can belong to one or more than one category.

Secreted Protein-Related Sequences

[007]   Secreted proteins, also referred to as secreted factors, include proteins that are produced by cells and exported extracellularly, extracellular fragments of

transmembrane proteins that are proteolytically cleaved, and extracellular fragments of cell surface receptors, which fragments may be soluble. An example of a secreted protein is keratinocyte growth factor (KGF), which stimulates the growth of keratinocytes, and is useful for repairing tissue after chemotherapy or radiotherapy.

[008]   Many and widely variant biological functions are mediated by a wide variety of different types of secreted proteins. Yet, despite the sequencing of the human genome, relatively few pharmaceutically useful secreted proteins have been identified. It would be advantageous to discover novel secreted proteins or polypeptides, and their corresponding polynucleotides that have medical utility.

[009]   Pharmaceutically useful secreted proteins of the present invention will have in common the ability to act as ligands for binding to receptors on cell surfaces in ligand/receptor interactions, to trigger certain intracellular responses, such as inducing signal transduction to activate cells or inhibit cellular activity, to induce cellular growth, proliferation, or differentiation, or to induce the production of other factors that, in turn, mediate such activities.

[010]   The cell types having cell surface receptors responsive to secreted proteins are various, including, for example, stem cells; progenitor cells; and precursor cells and mature cells of the hematopoietic, hepatic, neural, lung, heart, thymic, splenic, epithelial, pancreatic, adipose, gastrointestinal, colonic, optic, olfactory, bone and musculoskeletal lineages. Further, the hematopoietic cells can be red blood cells or white blood cells, including cells of the B lymphocytic (B cell), T lymphocytic (T cell), dendritic, megakaryocytic, natural killer (NK), macrophagic, eosinophilic, and basophilic lineages. The cell types responsive to secreted proteins also include normal cells or cells implicated in disorders or other pathological conditions.

[011]   As an example, certain of the secreted proteins of the present invention can stimulate T or B cell growth or differentiation by interacting with precursor T or B cells or hematopoietic progenitor cells, or bone marrow stem cells. As another example, certain secreted proteins of the present invention can maintain stem cells, progenitor cells or precursor cells in an undifferentiated state. As a further example, certain secreted proteins of the present invention can regulate bone growth by stimulation or inhibition thereof, secretion of insulin, glucose metabolism, cell proliferation, response to microbial infection, and regeneration of tissues including

neural, muscular, and epithelial. Moreover, certain secreted proteins of the present invention can induce apoptosis such as in cancer cells or inflammatory cells.

[012] Certain of the secreted proteins of the present invention are useful for diagnosis, prophylaxis, or treatment of disorders in subjects that are deficient in such secreted proteins or require regeneration of certain tissues, the proliferation of which is dependent on such secreted proteins, or requires an inhibition or activation of growth that is dependent on such secreted proteins. Examples of such disorders include cancer, such as bone cancer, brain tumors, breast and ovarian cancer, Burkitt's lymphoma, chronic myeloid leukemia, colon cancer, endocrine system cancers, gastrointestinal cancers, gynecological cancers, head and neck cancers, leukemia, lung cancer, lymphomas, malignant melanoma, metastases, multiple endocrine neoplasia, myelomas, neurofibromatosis, pancreatic cancer, pediatric cancers, penile cancer, prostate cancer, disorders related to the Ras oncogene, retinoblastoma (RB), sarcomas, skin cancers, testicular cancer, thyroid cancer, urinary tract cancers, and von Hippel-Lindau syndrome.

[013] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of hematopoeisis, including thrombosis; bleeding; anemias, e.g., iron deficiency and other hypoproliferative anemias, megaloblastic anemias, hemolytic anemias, acute blood loss, and aplastic anemia; hemoglobinopathies; disorders of granulocytes and monocytes; myelodysplasias and related bone marrow failure syndromes; polycythemias, e.g., polycythemia vera; acute and chronic myeloid leukemia, and other myeloproliferative diseases, e.g., malignancies of lymphoid cells; stimulation of replacement cell growth following irradiation or chemotherapy; and plasma cell disorders.

[014] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of hemostasis, such as disorders of the platelet and vessel wall, disorders of coagulation and thrombosis, and anticoagulant, fibrinolytic and antiplatelet therapies.

[015] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the cardiovascular system including disorders of the heart, such as heart failure; congenital heart disease; rheumatic fever; cor pulmonale; cardiomyopathies e.g., myocarditis; pericardial disease; cardiac tumors; cardiac manifestations of systemic diseases; and vascular diseases, such as

acute myocardial infarction, ischemic heart disease, hypertensive vascular disease, diseases of the aorta, and vascular diseases of the extremities.

[016]   Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the respiratory system, such as asthma, hypersensitivity pneumonitis, e.g., with pulmonary infiltration, pneumonia, necrotizing pulmonary infections, bronchiectasis, cystic fibrosis, chronic bronchitis, emphysema and airway obstruction, interstitial lung diseases, primary pulmonary hypertension, pulmonary thromboembolism, disorders of the pleura, mediastinum, and diaphragm, disorders of ventilation, sleep apnea, and acute respiratory distress syndrome.

[017]   Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the kidney and urinary tract, such as, for example, chronic renal failure and glomerulopathies.

[018]   Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the gastrointestinal system, including disorders of the alimentary tract, such as, for example, peptic ulcer disease and related disorders, inflammatory bowel disease, irritable bowel syndrome; disorders of the liver and biliary tract, such as, for example, hyperbilirubinemias, acute viral hepatitis, chronic hepatitis, and cirrhosis; and disorders of the pancreas, such as acute or chronic pancreatitis.

[019]   Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the immune system, connective tissue, and joints, including, for example, autoimmune diseases, primary immune deficiency diseases, human immunodeficiency virus diseases, allergies, systemic lupus erythematosus, rheumatoid arthritis, systemic sclerosis, Sjogren's syndrome, ankylosing spondylitis, reactive arthritis, vasculitis, sarcoidosis, amyloidosis, osteoarthritis, gout, psoriatic, and other arthritis.

[020]   Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the endocrine system, including, for example, disorders of the pituitary, hypothalamus, neurohypophysis, thyroid gland, adrenal cortex, testes, ovary, and other organs of the female reproductive system, such as breast; as well as pheochromocytoma, diabetes mellitus, and hypoglycemia.

[021]   Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of bone and mineral metabolism, and other

metabolic processes, including, for example, diseases of the parathyroid gland and other hyper- and hypocalcemic disorders, osteoporosis, Paget's disease and other dysplasia of bone, disorders of lipoprotein metabolism, hemochromatosis, porphyries, disorders of purine and pyrimidine metabolism, Wilson's disease, lysosomal storage diseases, glycogen storage diseases, lipodystrophies, and other primary disorders of adipose tissue.

[022]   Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the central nervous system, including, for example, seizures and epilepsy, cerebrovascular diseases, Alzheimer's disease and other extrapyramidal disorders, ataxic disorders, amylotrophic lateral sclerosis and other motor neuron diseases, disorders of the autonomic nervous system, diseases of the spinal cord, including spinal cord injury, primary and metastatic tumors of the nervous system, multiple sclerosis, and other demyelinating diseases, as well as chronic and recurrent meningitis.

[023]   Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of nerves or muscle, including, for example, Guillain-Barre Syndrome, myasthenia gravis and other diseases of the neuromuscular junction, polymyositis, dermatomyositis, muscular dystrophies, and other muscle diseases.

[024]   Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the skin, including, for example, eczema, psoriasis, cutaneous infections, acne, and other common skin disorders, and immunologically mediated skin diseases.

[025]   The agonists or antagonists of the secreted proteins herein or fragments thereof can be useful in treating elevated levels of such proteins in ny of the disorders above, and including angina, anoxia, arrhythmias, asthma, atherosclerosis, benign prostatic hyperplasia, Buerger's Disease, cardiac arrest, cardiogenic shock, cerebral trauma, Crohn's Disease, congenital heart disease, mild congestive heart failure (CHF), severe congestive heart failure, cerebral ischemia, cerebral infarction, cerebral vasospasm, cirrhosis, diabetes, dilated cardiomyopathy, endotoxic shock, gastric mucosal damage, glaucoma, head injury, hemodialysis, hemorrhagic shock, hypertension (essential), hypertension (malignant), hypertension (pulmonary), hypertension (e.g., pulmonary, after bypass), hypoglycemia, inflammatory arthritis, ischemic bowel disease, ischemic disease, male penile erectile dysfunction, malignant

hemangioendothelioma, myocardial infarction, myocardial ischemia, prenatal asphyxia, postoperative cardiac surgery, prostate cancer, preeclampsia, Raynaud's Phenomenon, renal failure (acute), renal failure (chronic), renal ischemia, restenosis, sepsis syndrome, subarachnoid hemorrhage (acute), surgical operations, status epilepticus, stroke (thromboembolic), stroke (hemorrhagic), Takayasu's arteritis, ulcerative colitis, uremia after hemodialysis, and uremia before hemodialysis.

[026]   Secreted proteins can be screened for functional activities in appropriate functional assays, as is conventional in the art. Such assays include, for example, *in vitro* and *in vivo* assays for factors that stimulate the proliferation or differentiation of stem cells, progenitor cells, or precursor cells into T cells, B cells, pancreatic islet cells, bone cells, neuronal cells, etc.

[027]   The tetratricopeptide repeat (TPR) is an example of a protein domain characteristic of a protein family, and is present in some of the secreted polypeptides of the invention. The TPR family is characterized by a degenerate 34 amino acid sequence present in a wide variety of proteins; it mediates protein-protein interactions, and is involved in scaffold formation and the assembly of multiprotein complexes (http://pfam.wustl.edu/cgi-bin/getdesc?name=TPR). Secreted protein-related sequences can also possess or interact with cytochrome P450 domains, which are involved in the oxidative degradation of various compounds, including environmental toxins and mutagens (http://pfam.wustl.edu/cgi-bin/getdesc?name=p450). Secreted protein-related sequences, e.g., cholesteryl ester transfer protein and phospholipid transfer protein, can also possess or interact with the LBP/BPI/CETP domain, which is characteristically found in lipid-binding serum glycoproteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=LBP_BPI_CETP). Secreted protein-related sequences can also possess or interact with peptidase S8 domains, also known as subtilase domains, which are comprised of serine proteases with a wide range of peptidase activities, including exopeptidase, endopeptidase, oligopeptidase, and omega-peptidase activity (http://pfam.wustl.edu/cgi-bin/getdesc?name=Peptidase_S8).. Secreted protein-related sequences can also possess or interact with adh_short, or short-chain dehydrogenase domains, which are found in a large family of proteins, and are made up of short-chain dehydrogenases and reductase enzymes; most family members function as NAD- or NADP- dependent oxidoreductases (http://pfam.wustl.edu/cgi-bin/getdesc?name=adh_short).

[028]   The inventors herein have identified novel secreted proteins using an algorithm that is constructed on the basis of a number of attributes including hydrophobicity, two-dimensional structure, prediction of signal sequence cleavage site, and other parameters. Based on such algorithm, a sequence that has a secreted tree vote of 0.5 – 1.0, preferably, 0.6 – 1.0, is believed to be a secreted protein.

Transmembrane Protein-Related Sequences

[029]   Transmembrane proteins extend into or through the cell membrane's lipid bilayer; they can span the membrane once, or more than once. Transmembrane proteins that span the membrane once are "single transmembrane proteins" (STM), and transmembrane proteins that span the membrane more than once are "multiple transmembrane proteins" (MTM). Examples of transmembrane proteins include the insulin receptor, adenylate cyclase, and intestinal brush border esterase.

[030]   A single transmembrane protein typically has one transmembrane (TM) domain, spanning a series of consecutive amino acid residues, numbered on the basis of distance from the N-terminus, with the first amino acid residue at the N-terminus as number 1. A multi-transmembrane protein typically has more than one TM domain, each spanning a series of consecutive amino acid residues, numbered in the same way as the STM protein.

[031]   Transmembrane proteins, having part of their molecules on either side of the bilayers, have many and widely variant biological functions. They transport molecules, e.g., ions or proteins across membranes, transduce signals across membranes, act as receptors, and function as antigens. Transmembrane proteins are often involved in cell signaling events; they can comprise signaling molecules, or can interact with signaling molecules. For example, tyrosine kinases can be transmembrane receptor proteins. Abnormalities of receptor tyrosine kinases are associated with human cancers; tumor cells are known to use receptor tyrosine kinases in transduction pathways to achieve tumor growth, angiogenesis and metastasis. Therefore, receptor tyrosine kinases represent pivotal targets in cancer therapy. It would be similarly advantageous to discover novel transmembrane proteins or polypeptides, and their corresponding polynucleotides that have additional medical utility.

[032]   The transmembrane polypeptides of the invention, like the secreted polypeptides, also have many different functional domains, and belong to a wide variety of Pfam families. Transmembrane protein-related sequences can possess or

11

interact with immunoglobulin (ig) domains, which are characteristically found in the immunoglobulin superfamily, comprised of hundreds of proteins, with various functions (http://pfam.wustl.edu/cgi-bin/getdesc?name=ig). Transmembrane protein-related sequences can also possess or interact with ion_trans domains, which are polypeptides characterized by six transmembrane helices, and which transport ions across membranes (http://pfam.wustl.edu/cgi-bin/getdesc?name=ion_trans). Proteins in this family can demonstrate specificity for particular ions, e.g., sodium, potassium, and calcium. Transmembrane protein-related sequences can also possess or interact with integrase core domains, which mediate the integration of a DNA copy of a viral genome into a host chromosome; e.g., HIV integrase catalyses the incorporation of virally derived DNA into the human genome, presenting a target for the development of new therapeutics for the treatment of AIDS (http://pfam.wustl.edu/cgi-bin/getdesc?name=rve). Transmembrane protein-related sequences can also possess or interact with domains designated as differentially expressed in neoplastic vs. normal cells "DENN" domains, which are involved in signal transduction. Characteristically, these domains are found in protein components of signaling pathways that utilize rab proteins or mitogen-activated protein (MAP) kinases (http://pfam.wustl.edu/cgi-bin/getdesc?name=DENN).

[033] Transmembrane protein-related sequences can also possess or interact with acyl coA binding protein (ACBP) domains, which are protein domains that bind medium- and long-chain acyl-CoA esters with high affinity (http://pfam.wustl.edu/cgi-bin/getdesc?name=ACBP). Membrane-related sequences also possess or interact with SPFH domain/band 7 family (Band_7) domain, which are protein domains that include a transmembrane segment, and regulate cation conductivity (http://pfam.wustl.edu/cgi-bin/getdesc?name=Band_7).

[034] Transmembrane proteins that are differentially expressed on the surface of cancer cells, particularly those that are differentially expressed on the surface of cancer cells but not on the surface of normal tissues, such as heart and lung, are desirable targets for production of antibodies, e.g., diagnostic antibodies or therapeutic antibodies, such as antibodies that mediate ADCC or CDC to effect tumor cell killing.

[035] Transmembrane proteins with extracellular fragments that can be cleaved can be useful as secreted proteins to effect ligand/receptor binding so as to mediate intracellular responses, such as signal transduction. Transmembrane proteins

that act as receptors, and possess a ligand binding extracellular portion exposed on a cell surface and an intracellular portion that interacts with other cellular components upon activation can be also be useful as transmembrane proteins to mediate intracellular responses, such as signal transduction.

Kinase-Related Sequences

[036]   A kinase is an enzyme that catalyzes the transfer of phosphate groups from phosphate donors to acceptor substrates. Kinase substrates include, but are not limited to, proteins and lipids. Sequences of the invention that phosphorylate protein substrates are designated "Pkinases." Examples of kinase-related sequences include calcium, calmodulin-dependent protein kinase II, myosin light chain kinase, and phosphatidlyinositol kinase.

[037]   Kinases and phosphatases are counteracting: kinases add phosphate groups and phosphatases liberate phosphate groups. The counteracting activities of kinases and phosphatases provide cells with a "switch" that can turn on or turn off the function of various proteins. The activity of any protein regulated by phosphorylation depends on the balance, at any given time, between the activities of the kinase(s) that phosphorylate it, and the phosphatase(s) that dephosphorylate it. Phosphorylation plays a important role in intercellular communication during development, homeostasis, and the function of major bodily systems, including the immune system.

[038]   In conjunction with phosphatases, kinases control such diverse and essential cellular processes as transcription, cell division, cell cycle progression, differentiation, cytoskeletal function, apoptosis, receptor function, learning and memory, hematopoeisis, fertilization, neural transmission, muscle contraction, non-muscle motor function, glycogen metabolism, and hormone secretion.

[039]   Most kinases act within a network of kinases and other signaling effectors, and are modulated by autophosphorylation and phosphorylation by other kinases (Manning et al., 2002). Intracellular signaling involves a multitude of diverse mechanisms that combine to modulate the activity of individual proteins in response to different biological inputs.

[040]   Defects in cell signal transduction pathways are responsible for a number of disorders, including the majority of cancers, immune disorders, and many inflammatory conditions, including, but not limited to, Crohn's disease (Geffen and Man, 2002; Van Den Blink et al., 2002; Lodish 1999). Over-expression and/or structural alteration of kinases, for example, receptor tyrosine kinase family members,

is often associated with human cancers. For example, tumor cells are known to use receptor tyrosine kinases in transduction pathways to achieve tumor growth, angiogenesis and metastasis. Therefore, receptor tyrosine kinases represent pivotal targets in cancer therapy. A number of small molecule receptor tyrosine kinase inhibitors have been synthesized, are in clinical trials, are being analyzed in animal models, or have been marketed. Inhibitory mechanisms include ligand-dependent down regulation, e.g., by the adaptor Cbl (Brunelleschi et al., 2002).

[041]   Kinase-related sequences can possess or interact with protein kinase (pkinase) domains, which share a conserved catalytic core common in serine/threonine and tyrosine protein kinases (http://pfam.wustl.edu/cgi-bin/getdesc?name=pkinase). Kinase-related sequences can also possess or interact with A-kinase anchoring protein 95 (AKAP95) domains, which comprise two zinc fingers, and have been implicated in chromosome condensation (http://pfam.wustl. edu/cgi-bin/getdesc?name=AKAP95). Kinase-related sequences can also possess or interact with inositol 1,3,4,-trisphosphate 5/6 kinase (Ins134_P3_kin) domains, which mediate the function of inositol 1.3.4-trisphosphate, a branch point in inositol phosphate metabolism (http://pfam.wustl.edu/cgi-bin/getdesc?name= Ins134_P3_kin).

[042]   Kinases, by virtue of their participation in many and varied intracellular activities, are useful as targets of therapeutic intervention such as, for example, in cancer and inflammation. Cells transfected with cDNA encoding a kinase can be used in screening for small molecule agonists or antagonists, for example.

Ligase-Related Sequences

[043]   Ligases are enzymes that join together, or ligate, two molecules. Ligase substrates include nucleic acids and proteins. For example, DNA ligases link two DNA molecules together; they play a role in DNA repair and replication. DNA ligases also are involved in the rearrangement of immunoglobulin gene segments, such as those responsible for the generation of antibody diversity. Examples of protein ligases include ubiquitin protein ligases, which add an ubiquitin molecule to an amino acid residue, typically as part of a peptide or polypeptide. Examples of nucleic acid ligases include DNA ligase I, DNA ligase III alpha, and T4 RNA ligase 2.

[044]   Ligases are also involved in cellular regulatory processes. For example, glutamate-cysteine ligase (GCL) is the first and rate-limiting enzyme involved in the biosynthesis of glutathione. Polymorphisms of human GCL account

14

for differences in sensitivity to environmental toxicants and chemotherapeutic agents in human cancer cell lines (Walsh et al., 2001). Also by way of example, glutamate-ammonia ligase, or glutamine synthetase (GS), is expressed at a higher than normal level in human primary liver cancer, and may be involved in hepatocyte transformation (Christa et al., 1994).

[045]   Ligase-related sequences can possess or interact with ATP dependent DNA ligase (DNA_ligase) domains, which can join two DNA fragments by catalyzing the formation of an internucleotide ester bond between a phosphate and a deoxyribose (http://pfam.wustl. edu/cgi-bin/getdesc?name= DNA_ligase). Ligase-related sequences can also possess or interact with glutamate-cysteine ligase (GCS) domains, which catalyze the rate-limiting step in the biosynthesis of glutathione. (http://pfam.wustl.edu/cgi-bin/getdesc?name=GCS). Ligase-related sequences can also possess or interact with 2',5' RNA ligase (2_5_ligase) domains, which ligate tRNA half molecules containing 2',3'-cyclic phosphate and 5' hydroxyl terminal to products containing a 2'5' phosphodiester linkage (http://pfam.wustl.edu/cgi-bin/getdesc?name=2_5_ligase).

[046]   Like kinases, ligases are also useful as targets for identification of agonists and antagonists, such as small molecule drugs.

Receptor-Related Sequences (Including Nuclear Hormone and T-Cell Receptors)

[047]   A receptor is a polypeptide that binds to a specific signaling molecule and initiates a cellular response. Receptors can be present on the cell surface or inside the cell. Example of receptor types include G-protein-linked receptors, ion channel-linked receptors, enzyme-linked receptors, T-cell receptors, thyroid hormone receptors, retinoid receptors, nuclear hormone receptors, and the related category of steroid hormone receptors, e.g., cortisol receptors (Alberts et al., 1994).

[048]   G-protein-linked receptors transduce extracellular signals into intracellular responses by interacting with guanine nucleotide binding proteins. The same ligand can activate many different G-protein-linked receptors. G-protein-linked receptors mediate cellular responses to a diverse range of signaling molecules, including hormones, neurotransmitters, and local mediators, which are varied in structure and function, and encompass proteins and small peptides, as well as amino acids and their derivatives, and fatty acids and their derivatives. Many signaling molecules are active at low concentrations, and their receptors often bind with high

affinity. Examples of G-protein-linked receptors include, but are not limited to, rhodopsins, olfactory receptors, and β-adrenergic receptors.

[049]   Ion channel-linked receptors are involved in synaptic signaling. These receptors regulate ion channels, to which they are linked. Some respond to signals from neurotransmitters, e.g., acetylcholine, serotonin, GABA, and glycine. A common mechanism of action for ion channel-linked receptors is to transiently open or close their respective ion channel, transiently changing the permeability of the membrane in which they reside to a specific ion or ions.

[050]   Enzyme-linked receptors can be linked to enzymes or can function as enzymes. Their ligand binding site is commonly on one side of the membrane, e.g., an extracellular domain, and the catalytic site is on the other, e.g., a cytoplasmic domain. Transmembrane tyrosine-specific protein kinase receptors for growth and differentiation factors are enzyme-linked receptors; examples include receptors for epidermal growth factor (EGF), platelet-derived growth factor (PDGF), fibroblast growth factors (FGFs), hepatocyte growth factors (HGF), insulin, insulin like growth factor-1 (IGF-1), nerve growth factor (NGF), vascular endothelial growth factor (VEGF), and macrophage colony stimulating factor (M-CSF).

[051]   Nuclear hormone receptors generally function by crossing the plasma membrane of target cells and binding to intracellular protein ligands. Ligand binding activates these receptors in some instances, exposing a DNA binding domain which regulates the transcription of specific genes. Generally, nuclear hormone receptors bind to specific DNA sequences adjacent to or in the vicinity of the genes regulated by their ligand. A host of cell type-specific regulatory proteins can collaborate with the nuclear hormone receptor to influence the transcription of specific genes or sets of genes (Alberts et al., 1994). Examples of nuclear hormone receptors include estrogen-related receptors, such as hERR1, which modulates the estrogen receptor-mediated response of the lactoferrin gene promoter (Yang et al., 1996), and is a transcriptional regulator of the human medium chain acyl coenzyme A dehydrogenase gene (Sladek et al., 1997). Examples of nuclear hormone receptors also include photoreceptor-specific nuclear receptors, such as NR2E3, which are part of a large family of nuclear receptor transcription factors involved in signaling pathways. NR2E3 plays a role in cone function and human retinal photoreceptor differentiation and degeneration (Milam et al., 2002; Kobayashi et al., 1999).

[052]   T-cell receptors are membrane proteins comprised of two disulfide-linked polypeptide chains, each with two immunoglobulin-like domains. They display a similarity to antibodies in that they have a variable amino-terminal region and a constant carboxyl-terminal region which is coded for by variable, joining, and constant region genes (Wei et al., 1997; Alberts et al., 1994). Rearrangement of T-cell receptor genes have been associated with human T-cell leukemias (Fisch et al., 1993).

[053]   Receptors are involved in cellular processes that regulate growth and differentiation. Their dysregulation can lead to hyperproliferative conditions, and they are common therapeutic targets. For example, the EGF receptor is aberrantly activated in neoplasia, especially in tumors of epithelial origin. EGF receptor antagonists can successfully treat some of these tumors, either alone or in combination with chemotherapy or ionizing radiation (Kari et al., 2003). The progesterone receptor, an intracellular steroid hormone receptor, plays a role in the development and function of the mammary gland, the uterus, and the ovary. Mutation or aberrant expression of the progesterone receptor, or its regulatory molecules, can affect its normal function and lead to cancer (Gao and Nawaz, 2002).

[054]   Receptors are also involved in cellular processes that regulate inflammation and immunity. For example, members of the type 1 interleukin-1 receptor family mediate immune and inflammatory responses, and function in host defense. (O'Neill, 2002). Their activation can lead to the activation of signaling cascades, e.g., pathways involving transcription factors and protein kinases, resulting in an inflammatory response (O'Neill, 2002). Another mechanism by which receptors regulate inflammation and immunity is by their selective expression, at discrete stages of differentiation, by cells involved in the inflammatory response. For example, expression of the triggering receptor expressed on myeloid cells (TREM-1) and the myeloid DAP12-associating lectin (MDL-1) are correlated with myelomonocytic differentiation. These receptors are more highly expressed in differentiated cells, are involved in monocyte activation and the inflammatory response, and are expressed at a lower level in malignant compared to normal cells (Gingras et al., 2002).

[055]   Receptor-related sequences can possess or interact with seven transmembrane receptor (7tm_1) domains, which are protein domains with a structural framework comprising seven transmembrane helices found in receptors, e.g., receptors in the rhodopsin family with a wide range of functions, activated by

17

ligands that vary widely in structure and character (http://pfam.wustl.edu/cgi-bin/getdesc?name=7tm_1). Receptor-related sequences can also possess or interact with L1 transposable element (transposase_22) domains, some of which have been characterized to exhibit reverse transcriptase activity, and some of which are capable of retrotransposition. Receptor-related sequences can also possess or interact with a SH2 domain, which is a protein domain of about 100 amino acid residues found in many intracellular signal-transducing proteins, that can regulate intracellular signaling cascades by interacting with phosphotyrosine-containing target peptides in a sequence-specific and phosphorylation-dependent manner (http://pfam.wustl.edu/cgi-bin/getdesc?name=SH2). Receptor-related sequences can also possess or interact with LDL receptor domains, e.g., the low-density lipoprotein receptor repeat class B (Ldl_recept_b) domain, which comprises a conserved YWTD motif in multiple tandem repeats (http://pfam.wustl.edu/ cgi-bin/getdesc?name=ldl_recept_b). Receptor-related sequences can also possess or interact with ribosomal L10 (Ribosomal_L10e) domains, which are protein domains commonly found in the large ribosomal subunit (http://pfam.wustl.edu/cgi-bin/getdesc?name=Ribosomal_L10e).

[056]   Receptor-related sequences can possess or interact with zinc finger C4 type domains, which are DNA binding domains of nuclear hormone receptors that share a conserved cysteine-rich region of approximately 65 amino acids and regulate such diverse biological processes as pattern formation, cellular differentiation, and homeostasis (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc? PF00105). Receptor-related sequences can also possess or interact with a ligand binding domain of nuclear hormone receptors (hormone_rec), which are helical domains involved in the regulation of eukaryotic gene expression, cellular proliferation, and differentiation in target tissues (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00104). Receptor-related sequences can also possess or interact with Mov34 domains, which are regulatory subunits of the proteasome found in some regulators of transcription factors (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc? PF01398). Receptor-related sequences can also possess or interact with immunoglobulin domains, which are described above.

[057]   Receptors, and fragments of receptors can be used as therapeutics. For example, a ligand-binding portion, an effector-binding portion, and a kinase or phosphatase domain or consensus sequence can comprise fragments that

can function as agonists or antagonists enhance or reduce, e.g., ligand binding to the natural receptors, or effector function by the natural receptors.

Phosphatase-Related Sequences

[058]   A phosphatase, as indicated above, is an enzyme that catalyses the hydrolysis of esters of phosphoric acid. Its substrates include, but are not limited to, nucleic acids, proteins, and lipids. Together with kinases, phosphatases are active in a broad range of cellular functions, including transcription, cell division, cell-cycle progression, intermediate cellular metabolism, glycogen metabolism, lipogenesis and lipolysis, maintenance of electrochemical gradients, neuronal function, immune responses, intracellular vesicular transport, cytoskeletal function, sperm motility, and skeletal, cardiac, and smooth muscle function (Oliver and Shenolikar, 1998).

[059]   Disruption in these functions may lead to disorders. For example, as noted above, phosphatases regulate pathways of cell growth and programmed cell death; disruptions in these pathways can lead to abnormal cell growth, such as that which occurs in cancer. Mutations in serine/threonine protein phosphatase 2A (PP2A), a multifunctional regulator of cell growth and function, are associated with the increased growth of tumor cells (Schonthal, 2001). The tumor suppressor "phosphatase and tensin-homology deleted on chromosome 10" (PTEN) gene encodes PIP$_3$, a lipid phosphatase that dephosphorylates phosphatidlyinositol, thus countering the action of the oncogenes PI$_3$-kinase and Akt, which promote cell survival. PTEN has been identified as a tumor suppressor; it is deleted in multiple types of advanced human cancers.

[060]   Also as noted above, phosphatases regulate pathways that control immune function. For example, the CD45 phosphotyrosine phosphatase is one of the most abundant glycoproteins expressed on immune cells, and regulates T-cell signaling and development (Alexander, 2000). In addition, the serine/threonine phosphatase calcineurin plays a central role in lymphocyte activation, among other important and wide-ranging cellular functions (Baksh and Burakoff, 2000). Certain compounds, specifically, cyclosporine and FK-506 (Tacrolimus), have been found to inhibit the phosphatase activity of calcineurin, thereby suppressing the production of IL-2 and other cytokines. In addition, these compounds have recently been found to block the JNK and p38 signaling pathways triggered by antigen recognition in T-cells. Finally, phosphatase inhibitors have proven to be valuable as immune suppressant

drugs, and those in the field believe that modulators of phosphatase activity promise to be important immunoregulatory compounds (Allison, 2000).

[061] Phosphatase-related sequences can possess or interact with protein phosphatase 2C (PP2C) domains, which display Mn^{++} or Mg^{++} dependent protein serine/threonine phosphatase activity (http://pfam.wustl.edu/cgi-bin/getdesc? name=PP2C). Phosphatase-related sequences can also possess or interact with protein-tyrosine phosphatase (Y_phosphatase) domains, which catalyze the removal of a phosphate group attached to a tyrosine residue (http://pfam.wustl.edu/cgi-bin/getdesc?name=Y_phosphatase). Phosphatase-related sequences can also possess or interact with protein phosphatase inhibitor 1/DARPP-32 (DARPP-32) domains, which inhibit protein phosphatases, and play a role in regulating neurotransmitter pathways, receptors, and ion channels (http://pfam.wustl.edu/cgi-bin/getdesc? name=DARPP-32).

[062] Like kinases, phosphatases can be used as targets for therapeutic intervention, in cell-free or cell-based assays, for example, in screening for drugs, including small molecule drugs.

Protease-Related Sequences

[063] Proteases, also known as endopeptidases, are enzymes that cleave polypeptide chains by hydrolyzing peptide bonds at positions within the amino acid chain. Different proteases recognize different polypeptide sequences. Endopeptidase substrate specificities vary from broad to narrow; for example, subtilisins are relatively non-specific, and can cleave polypeptide chains with a wide variety of amino acid sequences, whereas thrombin is more specific and can only cleave polypeptide chains with an arginine residue on the carboxyl side of the susceptible peptide bond and glycine on the amino side. Additional examples of protease-related sequences include collagenases, trypsin, and damage-induced neuronal endopeptidase (Kiryu-Seo et al., 2000).

[064] Proteases mediate the continuous remodeling of living tissues. For example, the extracellular matrix, a tissue skeleton that mediates communication among cells, and influences the structure and function of associated tissues and organs, is continuously remodeled. A strictly controlled balance is maintained between breakdown of the extracellular matrix by proteases and reconstruction of the extracellular matrix. This continued matrix remodeling is a dynamic process that shapes the structure and function of tissues and organs (Wojtowicz-Praga, 1999).

[065]   Defects in protease function are responsible for a number of disorders, including cancer and other hyperproliferative disorders. Proteases are involved in the pathogenesis of such disorders both by virtue of their involvement in programmed cell death and tumor invasion and metastasis (Los et al., 2003; Stetler-Stevenson et al., 1993). Detection of the presence or characteristics of proteases can be used to screen for and diagnose prostate cancer (Karanazanashvili and Abrahamsson, 2003). Proteases are also involved in the pathogenesis of inflammatory and arthritic diseases, such as pancreatitis, osteoarthritis, and rheumatoid arthritis (Pfutzer and Whitcomb, 2001; Martel-Pelleteir et al., 2001; Lerch and Gorelick, 2000).

[066]   Protease-related sequences possess or interact with a variety of different protease domains, including domains belonging to the cysteine protease family, the serine protease family, and the metalloproteinase family (http://pfam.wustl.edu/cgi-bin/text search?terms=endopeptidase&search_what= all&sections =DE&sections=CC&size=10).

Phosphodiesterase-Related Sequences

[067]   Phosphodiesterases are enzymes that cleave phosphodiester bonds, i.e., bonds formed by two hydroxyl groups in an ester linkage to the same phosphate group, such as those between adjacent RNA or DNA nucleotides. Phosphodiesterases are found in both soluble and membrane-associated forms. Most phosphodiesterases act within a network of signal transduction molecules and other signaling effectors, and are modulated by components of these pathways. Phosphodiesterases regulate the metabolism and synthesis of cyclic nucleotides in signal-transduction pathways. They hydrolyze cAMP and cGMP, molecules that play an important and widespread role in signal transduction. Phosphodiesterases also repair damage to nucleic acids. Some phosphodiesterases are regulated primarily by calcium and calmodulin, others are regulated primarily by cGMP. They differ in their sensitivity to individual inhibitors, but all share a homologous catalytic region (Siegel, et al., 1999).

[068]   Examples of phosphodiesterases include nucleotide pyrophosphatases (NPP) and plasma membrane glycoprotein PC-1, which are present in elevated levels in the fibroblasts of patients with Lowe's syndrome (Funakoshi et al., 1992). Another example of a phosphodiesterase is myomegalin-like protein, which is expressed at high levels in the nucleus and cytoplasm of heart and skeletal

muscle (Soejima et al., 2001). Phosphodiesterases have demonstrated promise in cancer chemotherapy, analgesia, the treatment of Parkinson's disease, and the treatment of learning and memory disorders (Weishaar, et al., 1985).

[069]   Phosphodiesterase-related sequences can possess or interact with type I phosphodiesterase/nucleotide pyrophosphatase (phosphodiest) domains, which catalyze the cleavage of phosphodiester and phosphosulfate bonds (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF01663). Phosphodiesterase-related sequences can also possess or interact with 3'5'-cyclic nucleotide phosphodiesterase (PDEase) domains, which are involved in signal transduction (http://www.sanger.ac. uk/cgi-bin/Pfam/getacc?PF00233).

[070]   Phosphodiesterases (PDEs) are also useful as targets for therapeutic intervention, for example, for identification of agonists or antagonists, such as in the screening of small molecule inhibitors. A well known PDE-5 inhibitor, sildenafil citrate (Viagra®) is used for treatment of erectile dysfunction (Brock, 2000). The mechanism of action involves inhibition of PDE-5 enzyme and resulting increase in cyclic guanosine monophosphate (cGMP) and smooth muscle relaxation in the penis (Rosen and McKenna, 2002). Such inhibitors may also find use for treatment of severe pulmonary arterial hypertension. (Ghofrani et al., 2003).

Kinesin-Related Sequences

[071]   Cells transport proteins and organelles in an orderly and regulated manner along cytoskeletal filaments. Molecular motor proteins, such as kinesins, can carry such cargo along the cytoskeletal filaments to specific destinations, in a highly regulated manner. Exemplary membrane-bound cargoes include mitochondria, lysosomes, endoplasmic reticulum, and axonal vesicles (Vale, 2003). Kinesins also transport nonmembranous cargo, such as mRNAs, tubulin monomers, and intermediate filaments (Vale, 2003).

[072]   Kinesins, e.g., KIF11, function in the cell division process (Miki et al., 2001). In the nucleus, kinesins are necessary to establish spindle bipolarity, position chromosomes on metaphase plates, and maintain forces in the spindle. Several members of the kinesin family are associated with the chromosomes, and are likely to perform a role in mitotic chromosome movement (Miki et al., 2001). For example, the C-terminal kinesin KIFC1 is involved in the processes of meiosis, mitosis, and karyogamy (Miki et al., 2001). The kinesin GAKIN binds to the human analog of the Drosophila Discs Large tumor suppressor protein (hDlg), a membrane

associated guanylate kinase (Hanada, 2000). GAKIN undergoes translocation in T-lymphocytes upon their cellular activation (Hanada, 2000). The GAKIN/hDlg complex is also hypothesized to play a role in cell division (Hanada, 2000). Thus, the kinesin GAKIN plays a role in cell proliferation and T-cell mediated immune function.

[073] Kinesin-mediated intracellular transport is also implicated in as a mechanism of tumorigenesis. For example, kinesin transports the tumor suppressor adenomatous polyposis colon protein (APC) (Jimbo et al., 2002). The APC gene is mutated in both sporadic and familial colorectal tumors. The APC protein interacts with the microtubule plus-end-directed kinesin proteins KIF3A and KIF3B through an association with the kinesin superfamily-associated protein 3 (KAP3). Normally, the APC tumor suppressor is transported to its correct intracellular location at the tips of membrane protrusions. Mutant APCs derived from cancer cells, however, are unable to undergo kinesin-mediated transport, and do not accumulate with normal efficiency in clusters in the membrane protrusions, and thereby can not function efficiently as tumor suppressors.

[074] In view of the connection to cancer, investigators have sought small molecules to inhibit specific molecular motors in cells, such as the mitotic kinesin Eg5/Ksp (Mayer, 1999). In addition, others have found small molecule inhibitors of Eg5/Kap with low nanomolar affinity have anti-tumor activity, and one such agent has entered clinical phase I trials (Vale, 2003).

[075] In another arena, it has been proposed that impairing motor-driven delivery of MHC peptide complexes to the surface of dendritic cells could provide immunomodulation. Additionally, inhibiting the cell surface delivery of cytotoxic granules in T cells could help provide immunosuppressive therapy (Vale, 2003).

[076] Kinesin-related sequences can possess or interact with kinesin motor (kinesin) domains, which hydrolyze ATP and bind to microtubules to produce a motor-active force that transports intracellular vesicles and organelles (http://pfam.wustl.edu/cgi-bin/getdesc?name=kinesin). Kinesin-related sequences can also possess or interact with kinesin-associated protein (KAP) domains, which are non-motive domains that form a complex with kinesin (http://pfam.wustl.edu/cgi-bin/getdesc?name=KAP). Kinesin-related sequences can also possess or interact with

MyTH4 domains, which are present in the tail of the motor ATPase proteins kinesin and myosin (http://pfam.wustl.edu/cgi-bin/getdesc?name=MyTH4).

[077]　　　Kinesins, like kinases, are useful as targets for therapeutic intervention, for example, in screening for small molecule inhibitors for the treatment of cancer.

Immunoglobulin-Related Sequences

[078]　　An immunoglobulin is an antibody molecule, and is typically composed of heavy and light chains, each of which have constant regions that display similarity with other immunoglobulin molecules and variable regions that convey specificity to particular antigens. Most immunoglobulins can be assigned to classes, e.g., IgG, IgM, IgA, IgE, and IgD, based on antigenic determinants in the heavy chain constant region; each class plays a different role in the immune response.

[079]　　Immunoglobulins are characterized by a structural motif, the immunoglobulin (ig) domain, which is approximately one hundred amino acids long, is involved in protein-protein and protein-ligand interactions, and includes a conserved intradomain disulfide bond (http://pfam.wustl.edu/cgi-bin/getdesc?name=ig). It is one of the most common domains found among all known proteins, and is present in hundreds of proteins with diverse functions. Proteins with the ig domain comprise the immunoglobulin superfamily; members include antibodies, T-cell receptors, major histocomptability proteins, the CD4, CD8, and CD28 co-receptors, most of the invariant polypeptide chains associated with B and T cell receptors, leukocyte F_c receptors, the giant muscle kinase titin, and receptor tyrosine kinases (Janeway et al., 2001; Alberts, et al., 1994).

[080]　　Polypeptides with immunoglobulin-like domains can be markers for specific types of tissues and tumors. For example, a 43-kDa protein membrane antigen with two immunoglobulin-like domains in its extracellular region is expressed in normal human colonic and small bowel epithelium and > 95% of human colon cancers, but absent from most other human tissues and tumor types (Heath et al., 1997).

[081]　　Polypeptides with immunoglobulin-like domains are also involved in inflammation. For example, myelin oligodendrocyte glycoprotein, a myelin-specific protein found in the central nervous system, specifically binds to and activates complement, an effector of the immune system, via its extracellular immunoglobulin-like domain. By virtue of providing the means for an interaction between myelin and

24

the complement component of the immune response, myelin oligodendrocyte. glycoprotein is a modulator of central nervous system inflammation and has been predicted by those in the field to be relevant to the pathogenesis of demyelinating diseases such as multiple sclerosis (Johns and Barnard, 1997).

[082]   Immunoglobulin-related sequences can also possess or interact with leucine-rich repeat domains, which are involved in protein-protein interactions, and are used in molecular recognition processes as diverse as signal transduction, cell adhesion, cell development, DNA repair and RNA processing (http://pfam.wustl.edu/cgi-bin/getdesc?name =LRRNT). Immunoglobulin-related sequences can also possess or interact with fibronectin type III repeat (fn3) domains (http://pfam.wustl.edu/cgi-bin/getdesc?name=fn3), which contain binding sites for DNA and heparin. Immunoglobulin-related sequences can also possess or interact with WASp Homology domain 1 (WH1), which can bind the metabotropic glutamate receptors mGluR1alpha and mGluR5 (http://pfam.wustl.edu/cgi-bin/getdesc? name=WH1).

Glycosylphosphatidylinositol Anchor-Related Sequences

[083]   Glycosylphosphatidylinositol (GPI) anchor proteins are synthesized as single membrane proteins; the transmembrane segment is cleaved away in the endoplasmic reticulum, where a GPI membrane anchor is added. The resulting protein is bound to the non-cytoplasmic, i.e., either extracellular or luminal, side of the membrane by the GPI anchor. GPI anchor proteins can be dissociated from the membrane by phosphatidylinositol-inositol-specific phospholipase C (Alberts et al., 1994). Examples of GPI-anchor proteins include prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin (Vainberg et al., 1998), and carboxypeptidase M, which is associated with the differentiation of monocytes to macrophages (Rehli et al., 1995).

[084]   GPI anchor protein-related sequences can possess or interact with KE2 domains, which may contain a DNA binding leucine zipper motif(http://www. sanger.ac.uk /cgi-bin/Pfam/getacc?PF01920). GPI anchor protein-related sequences can also possess or interact with zinc carboxypeptidase (Zn_carbOpept) domains, which include carboxypeptidase H regulatory domains and carboxypeptidase A digestive domains (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00246).

Other Polypeptide-Related Sequences

Activator-Related Sequences

[085]   An activator is a molecule or collection of molecules that positively modulates the activity of a regulatory protein, or that binds to DNA and regulates one or more genes by increasing the rate of transcription. Regulatory protein activators contribute to an increase in protein activity. Transcriptional activators provide a positive control over gene transcription; for example, they can sense the internal condition of the cell and bind to a sequence of DNA near a target promoter, resulting in the transcription of an appropriate gene. Examples of activator-related sequences include template-activating factors, bacterial catabolite activators, and the coenzyme thiamine pyrophosphatase. Activator-related sequences, e.g., factors that influence viral replication and transcription, can be encoded by oncogenes (Nagata et al., 1995).

[086]   Activator-related sequences can possess or interact with SH2 domains, which are protein domains of about 100 amino acid residues found in many signal-transducing proteins. SH2 domains can regulate signaling cascades, e.g., by interacting with phosphotyrosine-containing target peptides in a sequence-specific and phosphorylation-dependent manner (http://pfam.wustl.edu/cgi-bin/getdesc? name=SH2). Activator-related sequences also possess or interact with nucleosome assembly protein (NAP) domains, which regulate gene expression, and are accessible to histones (http://pfam.wustl.edu/cgi- bin/getdesc?name=NAP).

Adaptor-Related Sequences

[087]   Adaptors are proteins involved in the process of capturing specific cargo molecules into membrane-bound vesicles for transport through the cell. Different adaptors recognize different receptors for cargo molecules, and also recognize different vesicle coat proteins, accounting, in part, for the specificity of the content of intracellular vesicles bound to specific destinations within the cell (Kirsch et al., 1999). Examples of adaptor-related sequences include adaptins, clathrins, adaptor-related protein complex subunits, and Cas ligand with multiple Src homology 3 domains (CMS) adaptors.

[088]   Adaptor-related sequences can possess or interact with src homology 3 (SH3) domains, which are small protein modules of approximately 50 amino acid residues found in a variety of intracellular or membrane-associated proteins. SH3 domains are often indicative of a protein involved in signal

transduction events related to cytoskeletal organization. (http://pfam.wustl.edu/cgi-bin/getdesc?name=SH3). Adaptor-related sequences also possess or interact with the adaptin N-terminal (Adaptin_N) protein domain, which is found in the N terminal region of various adaptor protein complexes. The N-terminal region of adaptor proteins is relatively constant in comparison to the C-terminal (http://pfam.wustl.edu/cgi-bin/getdesc?name=Adaptin_N).

Adhesion Molecule-Related Sequences

[089]   Adhesion molecules are molecules that mediate the adhesion of cells with other cells, and with the extracellular matrix. Examples of adhesion molecules include members of the immunoglobulin superfamily, integrins, cadherins, selectins, and transmembrane proteoglycans. The adhesion molecule carcinoembryonic antigen (CEA) is present nearly exclusively on cancer cells, and is expressed on the cell surface of approximately 80% of all solid cancerous tumors (Berinstein et al., 2002).

[090]   Adhesion molecule-related sequences can possess or interact with the immunoglobulin (ig) domain, which are described above. Adhesion molecule-related sequences can also possess or interact with integrin alpha cytoplasmic region (integrin_A) domains, which comprise the short, intracellular region of the integrin alpha chain http://pfam.wustl.edu/cgi-bin/getdesc?name=integrin_A).

Antigen-Related Sequences

[091]   An antigen is a molecule that provokes an immune response; they include both foreign antigens and autoantigens. Antigens can be expressed in a tissue-specific manner and their expression can be developmentally regulated. For example, the heat stable antigen HSA is expressed in both a tissue-specific manner, i.e., it is restricted to hematopoeitic cells, and a developmentally-regulated manner, i.e., it is more highly expressed in immature precursor cells than in terminally differentiated cells (Wenger et al., 1993). Antigens can be expressed on the cell surface or inside the cell, e.g., in the nucleus or on intermediate filaments. Antigen-related sequences include sequences related to tumor antigens, which are expressed exclusively in tumor cells, or in greater amounts in tumor cells than in normal cells. Tumor antigens can be transmembrane proteins, with one or more transmembrane domains (Li et al., 1996; Linnenbach, et al., 1993).

[092]   Autoantigens, which are components of the body that provoke an immune response, are involved in the pathogenesis of autoimmune disease.

Autoantigens can be either selectively or ubiquitously expressed among cell and tissue types. They can be localized to any region of the cell, including the nucleus, nucleolus, nuclear envelope, and intermediate filaments (Racevskis et al., 1996). For example, pancreatic islet cell antigens are involved in the autoimmune pathogenesis of diabetes, and thyroid antigens are involved in autoimmune thyroid disease.

[093] Antigen-related sequences can possess or interact with the ICAp69 domain, which is characterized by a 69 kDa pancreatic islet cell autoantigen present in autoimmune (insulin-dependent) diabetes mellitus (http://pfam.wustl.edu/cgi-bin/getdesc?name=ICA69). Antigen-related sequences can also possess or interact with the Ku70/Ku80 C-terminal arm (Ku_C) or Ku70/Ku80 N-terminal alpha/beta (Ku_N) domains, which belong to the Ku family of peptides (http://pfam.wustl. edu/cgi-bin/getdesc?name=Ku_C; http://pfam.wustl. edu/cgi-bin/getdesc? name=Ku_N). Ku, an antigen associated with autoimmune disease, normally functions to bind DNA double-strand breaks and facilitate DNA repair, but induces autoimmunity under pathological conditions. Antigen-related sequences can also possess or interact with the bZIP transcription factor (bZIP) domain, which comprises a basic region and a leucine zipper region (http://pfam.wustl.edu/cgi-bin/getdesc? name=bZIP). Antigen-related sequences can possess or interact with YT521-B-like (YTH) domains, which comprise YT521-B, a tyrosine-phosphorylated nuclear protein domain that modulates alternative RNA splice site selection, and interacts with other nuclear proteins, e.g., scaffold attachment factor B, and Sam68, a 68-kDa substrate associated with Src during mitosis (http://pfam.wustl.edu/ cgi-bin/getdesc?name= YTH).

ATPase-Related Sequences

[094] ATPases are enzymes that use the energy of ATP hydrolysis to move ions or small molecules across a membrane against a chemical concentration gradient or electrical potential. For example, ATPases can maintain low intracellular calcium and sodium ion concentrations, and generate a low pH inside lysosomes, plant-cell vacuoles, and the lumen of the stomach. Vacuolar ATPases are ATP-dependent proton pumps that create pH gradients by transporting protons across membranes, while coupling the energy produced in the conversion of ATP to ADP with proton transport (Forgac, 1999). They can acidify or alkalinize cells, organelles, and extracellular compartments, and create voltage gradients that drive the secretion or absorption of ions and fluids (Wieczorek et al. 1999). Examples of ATPase-related

sequences include proton transporters, glucose transporters, multidrug resistance factors, calcium ATPases, and porins.

[095]   ATPase-related sequences can possess or interact with ATP synthase F/14-kDa subunit (ATP-synt-F) domains, which correspond to a 14-kDa subunit in the peripheral catalytic part of vacuolar ATPases (http://pfam.wustl.edu/ cgi-bin/getdesc?name=ATP-synt_F). ATPase-related sequences can also possess or interact with vacuolar (H^+)-ATPase C, D, G, and H subunit (V-ATPase) domains, which are membrane-attached sequences that generate an acidic environment (http://pfam.wustl.edu/cgi-bin/getdesc?name=V-ATPase_G).

ATP-Related Sequences

[096]   Adenosine trisphosphate (ATP) is a nucleotide comprising an adenine, a ribose, and a trisphosphate unit. The trisphosphate unit contains two phosphoanhydride bonds that confer an energy-rich property to ATP. The free energy liberated in the hydrolysis of one or both of these bonds can drive reactions that require an input of free energy. A wide range of physiological and pathological processes are driven by the energy of ATP, including cellular movement, the synthesis of biomolecules from precursors, muscle contraction, ciliary and flagellar function, intermediary metabolism, glycolysis, fatty acid oxidation, oxidative phosphorylation, and membrane transport (Ku et al., 1990). Examples of ATP-related sequences include ATPases, ATP synthases, ATP carrier proteins, and myosin.

[097]   ATP-related sequences can possess or interact with ATP-synthase subunit C protein domains (ATP-synt_C), which are protein domains that consist of two long terminal hydrophobic regions, and are implicated in the proton-conducting activity of ATPases (http://pfam.wustl.edu/cgi-bin/getdesc?name=ATP-synt_C). ATP-related sequences can also possess or interact with mitochondrial carrier protein (mito_carr) domains, which are involved in energy transfer across the inner mitochondrial membrane (http://pfam.wustl.edu/cgi-bin/getdesc? name= mito_carr).

Binding Protein-Related Sequences

[098]   A binding protein is a protein that binds to another molecule with specificity. Binding proteins can be involved in building macromolecular structures, e.g., in cytoskeletal assembly or scaffolding (Machesky et al., 1997). Proteins often exist in the cell in complexes with other proteins, nucleic acids, lipids, and/or small molecules. For example, steroid receptors, e.g., the progestin, estrogen, androgen,

and glucocorticoid receptors, bind to heat-shock proteins and FKBP52, a calcium-
regulated immunosuppressant, to form functional complexes (Peattie et al., 1992;
Sanchez et al., 1990). DNA binding proteins and general transcription factors bind to
the TATA box, a consensus sequence in a gene's promoter region that specifies the
position of transcription initiation, forming a functional transcription complex (Chalut
et al., 1995). Proteins can interact with multiple molecules simultaneously. For
example, Nedd4, an ubiquitin-protein ligase, can interact with multiple proteins and
lipids through its lipid binding domain and multiple protein binding domains (Jolliffe
et al., 2000).

[099]   Proteins utilize a large number of motifs to bind other molecules.
Binding protein-related sequences can possess or interact with the cold-shock DNA-
binding (CSD) domain, a conserved domain of about 70 amino acids that helps the
cell survive in temperatures below optimum growth temperature by inducing the
synthesis of proteins that negatively regulate transcription, translation, and
recombination, resulting in suppressed cell proliferation (http://pfam.wustl.edu/cgi-
bin/getdesc?name=CSD). Proteins induced by exposure to cold include DNA-binding
proteins, and cold inducible RNA binding proteins, which have RNA binding
domains at or near their N-termini (Nishiyama et al., 1997). For example, contrin, a
testis-specific DNA/RNA binding protein with a cold shock domain also has a large
number of phosphorylation sites, each of which can mediate intermolecular
interactions (Tekur et al., 1999). Contrin is involved in transcription of testis-specific
genes; its inactivation could provide a reversible male contraceptive.

[0100]   Binding protein-related sequences can possess or interact with the
ARID/BRIGHT DNA binding (ARID) domain, which is an approximately 100 amino
acid sequence involved in a wide range of DNA interactions, including, but not
limited to, interaction with AT-rich regions (http://pfam.wustl.edu/cgi-bin/getdesc?
name=ARID). ARID-encoding genes are involved in a variety of biological
processes, including regulation of cell growth, development, cell lineage gene
regulation, cell cycle control, and tissue-specific gene expression.

[0101]   Binding protein-related sequences can also possess or interact with
nucleosomal binding domains to facilitate binding within the nucleosome, a nuclear
structure comprised of chromosomal DNA and proteins. For example, the HMG14
and HMG17 (HMG14_17) domain is present in some nucleosome proteins, most
commonly, in proteins HMG14 and HMG17, members of a family designated as high

mobility group proteins, which form components of chromatin, and bind to nucleosomal DNA, regulating the interaction of the DNA with histone proteins (http://pfam.wustl.edu/cgi-bin/getdesc? name=HMG14_17).

[0102]   Binding protein-related sequences can also possess or interact with conserved motifs that recognize RNA, and allow the protein to bind RNA (http://pfam.wustl.edu/cgi-bin/textsearch?terms=rna+ binding&search_what= all&sections=DE&sections =CC&size=100). These motifs include the RNA recognition (rrm) domain, also known as a RRM, RBD, or RNP domain (http://pfam. wustl.edu/cgi-bin/getdesc?name=rrm). Numerous RNA binding proteins possess the rrm domain, including heterogeneous nuclear ribonucleoproteins (hnRNP) proteins, which are implicated in the regulation of alternative splicing, and LA proteins, which are among the main autoantigens in systemic lupus erythematosus (SLE).

[0103]   Binding protein-related sequences can also possess or interact with conserved motifs that mediate their binding to ions, e.g., calcium. Calcium-binding proteins such as calmodulin, the calcineurins, and their homologues and related proteins are widely used to regulate cellular processes (http://pfam.wustl.edu/cgi-bin/textsearch?terms=calcium +binding& search_what=all&sections=DE&sections= CC&size=100). Ion-binding proteins include phosphoproteins that bind to other molecules in an manner dependent on their phosphorylation state, and can regulate many types of molecules and processes, including those that utilize complex signaling cascades (Pang et al., 2001; Pang et al., 2002; Lin et al., 1999). Ion-binding protein-related sequences can possess or interact with the EF hand (efhand) domain, a calcium-binding domain that comprises a loop of twelve amino acids that coordinates a calcium ion in a pentagonal bipyramidal configuration and is flanked on both sides by a twelve amino acid alpha-helical domain (http://pfam.wustl. edu/cgi-bin/getdesc? name=efhand).

Breakpoint-Related Sequences

[0104]   A breakpoint is the location on a chromosome where a gene is disrupted, and one segment of the gene is severed from the other. Chromosomal breaks that disrupt coding or regulatory sequences can result in gene mutation. Chromosomal breaks can also serve as molecular landmarks, e.g., a break can be detected on Southern blots as the loss of an expected band and the appearance of two novel bands. Examples of breakpoint-related sequences include the sequences that generate the Philadelphia chromosome translocation, the sequences that generate the

31

chromosome translocation (t(1;7)(q42;p15)), which is implicated in Wilms' tumor, and the sequences that generate the chromosomal translocation t(18;21)(q22.1q21.3), which is implicated in Down syndrome.

[0105]   Breakpoints commonly occur in discrete regions of the chromosome. Breakage at these regions can lead to a recognized disease phenotype. One way of generating such a phenotype is by chromosomal translocation, i.e., chromosomes mutate by exchanging parts. When a segment from one chromosome is exchanged with a segment from another nonhomologous chromosome, two mutated chromosomes are simultaneously generated (Griffiths, et al., 1999). The Philadelphia chromosome, a mutation sometimes associated with chronic myelogenous leukemia (CML), is an example. It results from the translocation of a discrete segment of chromosome 22 into a discrete region of chromosome 9. Patients with the Philadelphia chromosome mutation generally have a better prognosis than CML patients with other characteristics.

[0106]   Acquired clonal chromosomal abnormalities are found in the malignant cells of most patients with leukemia, lymphoma, and solid tumors. Some of these abnormalities are the result of consistent chromosomal rearrangements. For example, in a preponderant number of chronic myelogenous leukemia cases, breakpoints at chromosome band 22q11 occur within a breakpoint cluster region of 5-6 kb (Weinstein et al., 1988).

[0107]   Chromosome rearrangements affecting band 3q21 are associated with a particularly poor prognosis in myeloid leukemia or myelodysplasia. These breakpoints cluster in a breakpoint cluster region of approximately 30 kb, located centromeric and downstream of the ribophorin I (RPN-I) gene (Weiser, 2002). The apoptotic gene *bcl-2*, was isolated as a breakpoint rearrangement in human follicular lymphomas and was shown to act as an oncogene that promoted cell survival rather than cell proliferation.

[0108]   Some proteins can act as leukemia or lymphoma-specific antigens for major histocompatibility complex-restricted T cell cytotoxicity. These include the breakpoint cluster region (bcr)-abl, and other fusion oncoproteins. Genetically engineered chimeric and humanized antibodies have demonstrated activity against overt lymphomas and leukemias. Radioimmunotherapy has produced significant therapeutic responses with minimal radiation exposure to normal tissues (Jurcic et al., 2000).

[0109] Breakpoint-related sequences can possess or interact with RhoGAP domains, also known as the breakpoint cluster region-homology domain, and mediates signal transduction by small G proteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=RhoGAP). Breakpoint-related sequences can also possess or interact with RhoGEF domains, which comprise approximately 200 amino acid residues that encode a guanine nucleotide exchange factor (http://pfam.wustl.edu/cgi-bin/getdesc?name=RhoGEF). Breakpoint-related sequences can also possess or interact with Plectin/S10 (S10_plectin) domains, which are found at the N-terminus of some isoforms of plectin and ribosomal S10 protein (http://pfam.wustl.edu/cgi-bin/getdesc?name=S10_plectin).

Carrier or Transport-Related Sequences

[0110] A membrane transport protein is an integral transmembrane protein that aids one or more molecules across a cell membrane. Most, if not all, types of molecules are transported across membranes, including proteins, ions, and fatty acids (Schaffer and Lodish, 1994). Even molecules such as water and urea, which can diffuse across pure phospholipid bilayers, are frequently accelerated by transport proteins. Transporters clear cells of toxins, and confer drug resistance on tumor lines (Ramalho-Santos et al., 2002). The rate of transport varies considerably among membrane transport proteins. Membrane transport proteins function in the plasma membrane and in intracellular organellar membranes, including the nuclear, mitochondrial, lysosomal, and vesicular membranes. For example, transportin, also known as karyopherin beta2, imports nuclear mRNA binding proteins from the cytoplasm across the nuclear membrane, into the nucleus (Bonifaci et al., 1997).

[0111] Membrane transport proteins can have either a broad or a narrow range of specificity for the transported substance. In mammalian cells, nucleoside transport across membranes is mediated by broad specificity transporters. Nucleoside transport plays a role in such diverse cellular functions as nucleotide synthesis, neurotransmission, and platelet aggregation. Nucleoside transporters carry chemotherapeutic nucleosides, and are a target of interest in chemotherapeutic and cardiac drug design (Griffiths et al., 1997; Ku et al., 1990).

[0112] Carriers are another class of membrane transport proteins; they bind to a solute and transport it across the membrane by undergoing a series of conformational changes. In contrast to channel proteins, transporters bind only one, or a few, substrate molecules at a time; after binding substrate molecules, they

undergo a conformational change such that the bound substrate molecules, and only those molecules, are transported across the membrane. Carriers transport a wide variety of molecules, including fatty acids across the plasma membrane (Schaffer and Lodish, 1994); purines, pyrimidines, and components of nucleosides across the nuclear membrane, and adenine nucleotides across the inner mitochondrial membrane (Battini et al., 1997).

[0113] Membrane transport-related sequences can possess or interact with vacuolar (H^+)-ATPase C, D, G, and H subunit (V-ATPase) domains, which are membrane-attached sequences that generate an acidic environment (http://pfam.wustl.edu/cgi-bin/getdesc? name=V-ATPase_C). Membrane transport-related sequences can also possess or interact with nucleoside transporter (nucleoside_tran) domains, which are found in proteins that transport nucleosides across the plasma membrane, and are employed to synthesize nucleotides via the salvage pathways in cells that lack their own de novo synthesis pathways (http://pfam.wustl.edu/cgi-bin/getdesc?name=Nucleoside_tran). Membrane transport-related sequences can also possess or interact with ATP synthase F/14-kDa subunit (ATP-synt-F) domains, which correspond to a 14-kDa subunit in the peripheral catalytic part of vacuolar ATPases (http://pfam.wustl.edu/cgi-bin/getdesc? name=ATP-synt_F). Membrane transport-related sequences can also possess or interact with mitochondrial carrier protein (mito_carr) domains, which are involved in energy transfer across the inner mitochondrial membrane (http://pfam.wustl.edu/cgi-bin/getdesc?name=mito_carr). Membrane transport-related sequences can also possess or interact with an AMP-binding enzyme (AMP-binding) domain, which is a domain rich in serine, threonine, and glycine, and is characterized by a conserved proline-lysine-glycine triplet sequence (http://pfam.wustl.edu/cgi-bin/getdesc?name=AMP-binding).

[0114] Membrane transport proteins, such as those expressed in cancer cells, are useful as targets for therapeutic intervention, for example, in the screening for small molecule inhibitors. Inhibition of membrane transport, as indicated above, may make cancer cells more susceptible to chemotherapy, for example.

Channel-Related Sequences

[0115]   Channel proteins transport water or specific types of ions down their concentration or electrical potential gradients. They form a protein-lined passageway across the membrane through which multiple water molecules or ions

move at a very rapid rate, e.g., up to 10^8 per second. The plasma membrane, for example, contains potassium-specific channel proteins that generate the cell's resting electric potential across the plasma membrane. Examples of channel-related sequences include the sodium hydrogen exchanger, sodium potassium ATPase, and the cystic fibrosis transmembrane regulator.

[0116]   Members of this subset of membrane transport proteins have wide-ranging functions in both normal physiology and in pathology. For example, the transport system that mediates the transmembrane exchange of sodium for hydrogen across the plasma membrane plays a physiological role in the regulation of intracellular pH, the control of cell growth and proliferation, stimulus-response coupling, metabolic responses to hormones, the regulation of cell volume, and the transepithelial absorption and secretion of several ions. The sodium-hydrogen exchanger also plays a role in cancer and in tissue and organ hypertrophy (Mahnensmith and Aronson, 1985).

[0117]   Channel-related sequences can possess or interact with sodium/hydrogen exchanger (Na_H_Exchanger) domains, which exchange sodium for hydrogen across a membrane in an electroneutral manner (http://pfam.wustl. edu/cgi-bin/getdesc? name=Na_H_Exchanger). Channel-related sequences can also possess or interact with neurotransmitter-gated ion-channel ligand binding (Neur_chan_LBD) domains, which form the extracellular domains of some ion channels (http://pfam.wustl.edu/cgi-bin/getdesc?name=Neur_chan_LBD). Channel-related sequences can also possess or interact with UBX domains, which are present in ubiquitin-regulatory proteins (http://pfam.wustl.edu/ cgi-bin/getdesc?name=UBX).

Checkpoint-Related Sequences

[0118]   The cell division cycle is the fundamental means by which living things are propagated. Fundamental to successful propagation is the faithful replication of DNA; a cell cycle control system exists to coordinate the cycle as a whole. The control system is regulated by brakes that can stop the cycle at specific checkpoints. Thus, the checkpoints arrest the cycle upon the occurrence of undesirable events, such as DNA damage, replication stress, or mitotic spindle disruption. For example, DNA lesions and disrupted replication forks are recognized by the DNA damage checkpoint and replication checkpoint, respectively. Checkpoints can also, for example, initiate protein kinase-based signal transduction cascades to activate downstream effectors that elicit cell cycle arrest, DNA repair, or

apoptosis. These actions prevent the conversion of aberrant DNA structures into inheritable mutations and minimize the survival of cells with unrepairable damage (Qin and Li, 2003).

[0119]   Dysregulation of the cell-cycle is a hallmark of tumor cells. Defective checkpoint function results in genetic modifications that contribute to tumorigenesis. Checkpoint function can be abrogated by many different mechanisms (Bast, et al., 2000). For example, cyclin-dependent kinases that normally are activated at a checkpoint can be inactivated or activated in an abnormal manner. Alternatively, the normal activities of the cyclin-dependent kinase inhibitors, phosphatases, or other regulatory molecules of the cell cycle can be altered. Tumor suppressors are among the classes of molecules that can effect cell cycle dysregulation. The abrogation of checkpoint function can alter the sensitivity of tumor cells to chemotherapeutics (Stewart et al, 2003).

[0120]   Checkpoint-related sequences can possess or interact with phosphoribosylaminoimidazole-succinocarboxamide synthase (SAICAR_synt) domains, which function in *de novo* purine synthesis (http://pfam.wustl.edu/cgi-bin/getdesc?name =SAICAR_synt). Checkpoint-related sequences can also possess or interact with WD40 domains, which comprise a domain of approximately 40 amino acids, which are sometimes present in tandem repeats (http://pfam.wustl.edu/cgi-bin/getdesc?name=WD40). Checkpoint-related sequences can also possess or interact with cyclin, C-terminal (cyclin_C) domains, which regulate cyclin dependent kinases (http://pfam.wustl.edu/cgi-bin/getdesc? name=cyclin_C). ·

[0121]   Thus, checkpoint related proteins, e.g., kinases, phosphatases, etc., are useful as targets for therapeutic intervention, such as in screening for small molecule drugs for the treatment of cancer, immune disorders, and inflammation.

Complex-Related Sequences

[0122]   Complexes are molecular entities comprised of two or more components. Molecular complexes within cells form functional units that carry out cellular operations. For example, complexes at the cell membrane perform structural and regulatory tasks, including regulating membrane traffic and maintaining organelle integrity. Complexes at the cytoskeleton perform static and dynamic roles with respect to cell shape, intracellular transport, and communication with the extracellular matrix. Complexes in the nucleus transcribe and regulate genes, and complexes at sites of protein synthesis translate and regulate proteins. Complexes can reside

36

intracellularly and/or extracellularly, e.g., in the extracellular matrix. Examples of complex-related sequences include cytoskeletal and filamentous proteins, ADP-ribosylation factor (ARF) proteins, and protein synthesis initiation factors (Amor et al., 1994).

[0123]   Complex-related sequences can possess or interact with ADP-ribosylation factor family (arf) domains, which are GTP-binding domains involved in protein trafficking (http://pfam.wustl.edu/cgi-bin/getdesc?name=arf). Complex-related sequences can also possess or interact with eukaryotic initiation factor domains, e.g., the eukaryotic initiation factor 4E (IF4E) domain, which recognizes and binds mRNA during protein synthesis (http://pfam.wustl.edu/cgi-bin/getdesc?name=IF4E). Complex-related sequences can also possess or interact with intermediate filament (filament) protein domains, which form filamentous structures typically 8 to 14 nm wide, and form components of the cytoskeleton and nuclear envelope, e.g., neurofilaments, cytokeratins, lamins, vimentin, and desmin (http://pfam.wustl.edu/cgi-bin/getdesc?name=filament).

Cytokine-Related Sequences

[0124]   A cytokine is an extracellular signaling protein or peptide that acts as a local mediator in communication among cells. Cytokines regulate proliferation and differentiation, for example, they mediate differentiation of cells in the hematopoeitic lineage. Examples of cytokines include interleukins, interferons, and colony stimulating factors of the hematopoeitic system. Some cytokines, e.g., interferons and interleukins, can be induced by viral activity, and possess antiviral activity (Sheppard et al., 2003). Cytokine-related sequences may enable the expression of a cytokine, for example, as a cytokine transcription factor (Kao et al., 1994). They can also be part of a cytokine effector pathway, for example, as an intracellular effector of cytokine-related cytoskeletal changes in response to events in the extracellular matrix (Hirsh et al., 2001; Joberty et al., 1999).

[0125]   Cytokine-related sequences can possess or interact with interferon-induced transmembrane protein (CD225) domains, which are associated with interferon-induced cell growth suppression (http://pfam.wustl.edu/ cgi-bin/getdesc?name=CD225). Cytokine-related sequences can also possess or interact with SelR (SelR) domains, which bind both selenium and zinc, and/or methionine sulfoxide reductase enzymatic domains (http://pfam. wustl.edu/cgi-bin/getdesc?name=SelR). Cytokine-related sequences can also possess or interact

with reverse transcriptase (rvt) domains, which are involved in RNA-directed DNA polymerase activity, an enzymatic activity that uses an RNA template to produce DNA for integration into a host genome (http://pfam.wustl.edu/cgi-bin/getdesc? name=rvt). Cytokine-related sequences can also possess or interact with L1 transposable element domains (Transposase_22), which are described above.

[0126] Cytokines, thus, are useful as therapeutic proteins for the treatment of disorders such as cancer, immune disorders, and inflammation.

Dehydrogenase-Related Sequences

[0127] Dehydrogenases are enzymes that catalyze the removal of hydrogen atoms in the absence of oxygen. They contribute to a wide range of enzymatic reactions, including those involved in amino acid degradation, amino acid synthesis, the citric acid cycle, fatty acid oxidation, fatty acid synthesis, glycolysis, the pentose phosphate pathway, photosynthesis, pyruvate oxidation, and oxidative phosphorylation (Walker et al., 1992). Examples of dehydrogenases include steroid dehydrogenases, NADH dehydrogenases, and glyceraldehyde-3-phosphate dehydrogenase.

[0128] Dehydrogenase-related sequences can possess or interact with glyceraldehyde 3-phosphate dehydrogenase, NAD binding (GPDH) domains, which play a role in glycolysis and gluconeogenesis by reversibly catalyzing the oxidation and phosphorylation of D-glyceraldehyde-3-phosphate to 1,3-diphospho-glycerate (http://pfam.wustl.edu/cgi-bin/getdesc?name=gpdh). Dehydrogenase-related sequences can also possess or interact with 3-hydroxyacyl-CoA dehydrogenase, NAD binding (3HCDH_N) domains, which catalyze the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA in fatty acid metabolism (http://pfam.wustl.edu/cgi-bin/getdesc? name=3HCDH_N).

Disease-Related Sequences

Amyotrophic Lateral Sclerosis

[0129] Amyotrophic Lateral Sclerosis (Lou Gehrig's Disease) is a neurodegenerative disease that affects the motor neurons. The disease displays multiple clinical variants and can affect motor neurons throughout the nervous system, e.g., the spinal cord and brainstem. One clinical variant, the autosomal recessive form of juvenile amyotrophic lateral sclerosis, has been mapped to the human chromosome 2q33-q34 region (Hadano et al., 2001). A protein family characterized by the HAP1 N-terminal conserved region (HAP1_N) domain possesses

a N-terminal conserved region from hypothetical protein products of ALS2CR3 genes found in the 2q33-2q34 region of chromosome 2 (http://pfam.wustl.edu/cgi-bin/getdesc?name= HAP1_N).

Gaucher's Disease

[0130] Gaucher's Disease is a genetic disease characterized by a deficiency of enzymes responsible for the breakdown and recycling of glycolipids, i.e., lipids with carbohydrate moieties, e.g., glucosylceramide; and sphingolipids, lipids with sphingosine moieties, e.g., sphingomyelin. Normally, the glycolipids and sphingolipids in the membranes of senescent cells are metabolized by a multi-step process that includes the activities of acid beta-glucosidases and saposins. When these activities are absent, or present in reduced amounts, glucosylceramide and sphingolipids accumulate, and produce the Gaucher's disease phenotype. The disease displays multiple clinical variants, and can manifest with central nervous system pathology, enlargement of organs, e.g., liver and spleen, and an increase in the level of the cytokine transforming growth factor beta (Zhao and Grabowski, 2002; Perez Calvo et al., 2000; Cormand et al., 1997). The variability in clinical presentation is consistent with the large number of different mutations observed in the acid beta-glucosidase and saposin genes.

[0131] Acid beta-glucosidases are enzymes that metabolize glycolipids. Saposins are small proteins that are described in more detail below. Mammalian saposins are synthesized as a single precursor molecule (prosaposin) with saposin-A (SAPA) and saposin-B (SapB_1; SapB_2) domains; prosaposin becomes an active saposin following a proteolytic activation reaction (http://pfam.wustl.edu/cgi-bin/getdesc?name=SAPA; http://pfam.wustl. edu/cgi-bin/getdesc?name=SapB_1; http://pfam.wustl.edu/cgi-bin/getdesc?name=SapB_1).

Huntington Disease

[0132] Huntington Disease is a progressive neurodegenerative genetic disorder characterized by dementia, psychiatric symptoms, and a choriform movement disorder. It is caused by an increased number of repeats of the codon CAG, which encodes the amino acid glutamine, in a gene located at the 4p16.3 region of chromosome 4, which codes for a protein called huntingtin. The polyglutamine tracts expressed by the mutant form of the gene selectively ablate striatal and cortical neurons, (Ho et al., 2001).

39

[0133] The Huntington Disease gene is widely expressed, but exerts tissue-specific effects on neurons (Lin et al., 1993). The gene expresses multiple distinct transcripts, and differential polyadenylation of the gene leads to the expression of transcripts of different sizes (Lin et al., 1993). There is a relative increase in the abundance of one transcript in the human brain, which has been hypothesized to account for the tissue-specific effects of the disease (Lin et al., 1993). The HAP1_N protein domain, described above, binds to the gene product, huntingtin, in a polyglutamine repeat-length-dependent manner (http://pfam.wustl.edu/cgi-bin/getdesc?name=HAP1_N). This domain is also found in several huntingtin-associated protein 1 (HAP1) homologues.

Multiple Sclerosis (MS)

[0134] Multiple sclerosis (MS) is a disease characterized by demyelination, i.e., the loss of the myelin coating, of nerve axons. Its clinical course varies among patients; these variations fall into two broad categories, a relapsing/remitting course, and a chronic progressive course. MS has a complex etiology; it has an autoimmune component, is influenced by genetics, and sometimes involves infectious agents. MS results from an abnormal immune response to one or more antigens present in the myelin sheaths that cover the nerve axons of genetically susceptible individuals, which may be preceded by exposure to a causal infectious agent (Oksenberg et al., 1999).

[0135] The genetic susceptibility to MS is determined by MS susceptibility genes, most of which demonstrate only a small to moderate effect on susceptibility, e.g., the major histocompatibility complex at chromosome 6p21 (Oksenberg et al., 1999). An etiological infectious agent has been isolated from the plasma and cerebrospinal fluid of patients with multiple sclerosis (Perron et al., 1997). This agent is a retroviral oncovirus, known as multiple sclerosis-associated retrovirus (MSRV), also called LM7, and is found in association with virions produced by the cultured cells of MS patients (Perron et al., 1997). MSRV proteins possess protein domains characteristic of retroviral proteins. These include the Gag P30 core shell protein (Gag_p30) domain, which is involved in viral assembly (http://pfam.wustl.edu/cgi-bin/getdesc?name=Gag_p30) and the reverse transcriptase (rvt) domain, which was described above.

Obesity

[0136] Although single-gene mutations have been shown to cause obesity in animal models, the most common forms of human obesity arise from the interactions of multiple genes, environmental factors, and behavior. Several genes have been shown to affect body weight regulation in humans and other animals. These include the ob, lep, CPE, ASIP, LEP, TUB, UPC, POMC, CCKAR, TNFA, and PPAR-γ genes (Comuzzie et al., 1998). Genetic regulation of body weight can be effected through diverse mechanisms. For example, the TUB gene family regulates body weight by encoding proteins that are phosphorylated in response to insulin, mediate insulin signaling, and are associated with a maturity onset obesity associated with insulin resistance (Ikeda et al., 2002). CCKAR genes regulate body weight in a different manner; they regulate the hormone cholecystokinin, which produces a feeling of satiety following food intake (Ritter et al., 1994).

[0137] Some genes that regulate body weight possess the WH1 domain, which is described above. Genes that regulate body weight can also possess or interact with the sprouty (sprouty) domain. This domain is found in sprouty proteins, which inhibit the Ras/mitogen-activated protein kinase cascade, a pathway initiated by receptor tyrosine kinases and involved in development (http://pfam.wustl.edu/cgi-bin/getdesc?name=Sprouty). Genes that regulate body weight can also possess or interact with a Tub (Tub) domain, which is found in Tubby, a mouse gene in which an autosomal recessive mutation resulting from a splicing defect causes maturity-onset obesity, insulin resistance and sensory deficits (http://pfam.wustl.edu/cgi-bin/getdesc?name=Tub).

Oncogene

[0138] An oncogene is any one of a large number of genes that can help make a cell cancerous. Typically, an oncogene is a mutant form of a normal gene, and is often a gene involved in the control of cell growth, division, or differentiation. Cells in higher organisms normally grow, divide, differentiate, and die under the regulation of other cells. Cancer cells proliferate, in part, because they are able to divide without input from other cells, as the result of accumulated mutations. Oncogenes include, but are not limited to, genes encoding GTP binding proteins, e.g., *ras*; growth factors, e.g., platelet-derived growth factor; growth factor receptors, e.g.,

platelet-derived growth factor receptor; kinases, e.g., *src*; nuclear proteins, e.g., *myc*; and tumor suppressors, e.g., retinoblastoma proteins.

[0139] The products of oncogenes are frequently proteins involved in cell signaling, e.g., kinases, GTP-binding proteins, and receptors. For example, many human cancers have a mutation in a ras gene (Alberts et al., 1994). The ras proteins belong to a large superfamily of monomeric GTPases, and relay signals from receptor tyrosine kinases to the nucleus, stimulating cell proliferation or differentiation. Ras proteins function as switches, cycling between an active state in which GTP is bound, and an inactive state, in which GDP is bound. A ras gene mutation can result in the translation of a protein that fails to hydrolyze its bound GTP, and persists abnormally in its active state, transmitting an intracellular signal for cell proliferation or differentiation even in the presence of regulatory non-proliferation and non-differentiation signals. Oncogene-related proteins can possess one of many ras protein domains (http://pfam.wustl.edu/cgi-bin/textsearch?terms=ras&search_what=all&sections=DE &sections=CC&size=100), including the sub-families Ras, Rab, Rac, Ral, Ran, Rap, and Ypt1. Oncogene-related proteins can also possess a Gtr1/RagA G-protein conserved region (gtr1_RagA) domain, which is found in some G-proteins of the Ras family, e.g., the RagA/B human homologues of the ras GTP binding protein Gtr1 (http://pfam.wustl.edu/cgi-bin/getdesc?name=Gtr1_RagA). Oncogene-related sequences can also possess or interact with an ATPase domain associated with diverse cellular activities; proteins with the AAA ('A'TPases 'A'ssociated with diverse cellular 'A'ctivities) domain can perform chaperone-like functions that assist in assembling, operating, or disassembling protein complexes. The domain includes a conserved region of approximately 220 amino acids that contains an ATP-binding site which can act as an ATP-dependent protein clamp to hold a protein in place (http://pfam.wustl.edu/cgi-bin/getdesc?name=AAA). Some oncogene-related sequences can also possess or interact with a C2 domain of approximately 116 amino-acid residues, which can be involved in calcium-dependent phospholipid binding and inositol-1,3,4,5-tetraphosphate binding, and is found, e.g., in some isozymes of protein kinase C (http://pfam.wustl.edu/cgi-bin/getdesc?name=C2). C2 domains are typically located between C1 domains (which bind phorbol esters and diacylglycerol) and protein kinase catalytic domains. Regions with homology to the C2 domain are present in many proteins, e.g., synaptotagmin.

Parkinson's Disease

[0140]   Parkinson's disease is a neurological disorder that affects movement control. Complex interactions among groups of nerve cells in the central nervous system coordinate to control movement. One such group of neurons is located in the substantia nigra of the midbrain; these neurons release the neurotransmitter dopamine, which allows an organism to fine-tune its movements. In Parkinson's disease, neurons of the substantia nigra progressively degenerate, leaving the patient with clinical symptoms that may include resting tremor, muscular rigidity, a slowness of spontaneous movement, and poor balance and motor coordination (Seigel et al., 1999).

[0141]   Parkinson's disease has multiple causes, including both genes and the environment. It also has multiple presentations, including juvenile-onset (before age 45) and adult onset (after age 45), and can be transmitted through either autosomal dominant or autosomal recessive mechanisms. In keeping with the diversity of etiologies, presentation, and genetic mechanisms, there are a large and diverse number of genes and gene products involved in the pathogenesis of Parkinson's disease. For example, the PARK2 gene, which encodes the protein parkin, is mutant in autosomal recessive juvenile parkinsonism. PARK2 is a ubiquitin protein ligase that is a component in the pathway that attaches ubiquitin to specific proteins, designating them for degradation (Fishman, and Oyler, 2002).

[0142]   Parkinson's disease-related sequences can possess or interact with synuclein domains, which are expressed on the cytoplasmic regions of proteins found predominantly in neurons (http://pfam.wustl.edu/cgi-bin/getdesc?name=Synuclein). Alpha-synuclein, which possesses a synuclein domain, is mutated in several families with autosomal dominant Parkinson's disease. Gamma-synuclein, which also possesses a synuclein domain, is overexpressed in breast and ovarian cancers (Lavedan, 1998).

Retinitis Pigmentosa

[0143]   Retinitis pigmentosa is a group of inherited retinopathies characterized by early stage loss of night vision, followed by loss of peripheral vision. Defects in any structural or functional proteins associated with the rod photoreceptor neurons of the retina, which are the cells that transduce light into a neuronal action potential, can lead to the disease (Seigel et al., 1999).

[0144]   GTPase regulators have been implicated in the pathology of retinitis pigmentosa. GTPase regulators are proteins that determine whether a GTP binding protein exists in a GTP-bound or GDP-bound state (Zhao et al., 2003); they are described in more detail below. GTPase regulators have a broad spectrum of intracellular functions, including intracellular vesicular transport. These proteins localize to a specific region of rod photoreceptor cells, in a narrow cilium that connects the cell body, where protein synthesis and basic metabolism takes place, with the rod outer segment, where light is transduced to an action potential of the optic nerve (Zhao et al., 2003). Proteins necessary for the light transduction process are made in the cell body and must be transported to the outer segment via vesicular transport mechanisms. Mutant GTPase regulators, which regulate vesicular transport, play a role in the pathogenesis of retinitis pigmentosa (Roepman et al., 2000). Retinitis pigmentosa-related sequences can possess or interact with a Tctex-1 domain, which is comprised of a dynein light chain, and can bind to the cytoplasmic tail of rhodopsins, which are light-sensing proteins present in retinal rod cells (http://pfam.wustl. edu/cgi-bin/getdesc?name=Tctex-1). Mutations in this domain that are responsible for retinitis pigmentosa inhibit this binding.

Alzheimer's Disease

[0145]   Alzheimer's disease is a neurodegenerative dementing illness. It is a genetically complex disease with multiple forms, including familial and sporadic forms, and early onset and late-onset forms. Mutations in at least four genes are known to cause Alzheimer's disease, and there is evidence for additional Alzheimer's loci (McKusick, 2003). One form of Alzheimer's disease is caused by mutations in the amyloid precursor gene, another form is associated with the apolipoprotein E4 allele, a third form is caused by a mutant presenilin-1 gene that encodes a seven-transmembrane domain protein, and a fourth form is caused by a mutant gene encoding a similar seven -transmembrane domain protein, presenilin-2 (McKusick, 2003).

[0146]   Consistent with its multiple etiologies, multiple clinical presentations, and multiple genetic loci, Alzheimer disease has a complex pathology. One facet of the pathology of Alzheimer's disease is the formation of amyloid plaques from amyloid precursor protein (Clark and Karlawish, 2003). Amyloid precursor protein can be processed *in vitro* by several different proteases such as secretases and caspases to yield peptide fragments, suggesting that these proteases may play a role in

the formation of pathogenic amyloid plaques *in vivo* (Suh and Checler, 2002).
Presenilins have been identified as likely candidates for the proteases that cleave
amyloid precursor protein to pathogenic peptide fragments *in vivo* (Selkoe, 2001).
Another facet of Alzheimer's disease pathology is an inflammatory component
mediated by microglial cells, the brain's primary immunoeffector cells (Tan et al.,
1999). Microglial cells are attracted to and activated by amyloid deposits; they
release inflammatory mediators that promote the aggregation of the deposits into
plaques, and also directly induce or promote neurodegeneration (Hoozemans et al.,
2002). Therefore, current treatment strategies include anti-inflammatory and
immunotherapeutic approaches, including vaccines (Weiner and Selkoe, 2002).

[0147] Alzheimer's disease-related sequences can possess or interact with
trypsin domains, which demonstrate a wide range of peptide degrading activities,
including exopeptidase, endopeptidase, oligopeptidase and omega-peptidase activities
(http://pfam. wustl.edu/cgi-bin/getdesc?name=trypsin). Alzheimer's disease-related
sequences can also possess or interact with low-density lipoprotein receptor (ldl_rece)
domains, which are characterized by seven successive cysteine-rich repeats of about
40 amino acids at the N-terminal region, and which are also present in receptors for
low density lipoprotein (LDL), the major cholesterol-carrying lipoprotein of plasma
(http://pfam.wustl.edu/cgi-bin/textsearch?terms=ldl_rece +&search_what=all&
sections =DE&sections=CC&size=100). Alzheimer's disease-related sequences can
also possess or interact with a PT repeat (pt_a) domain, which includes the
tetrapeptide XPTX, or a similar, conserved, sequence.

Williams-Beuren Syndrome

[0148] Williams-Beuren syndrome is a complex genetic developmental
disorder with multisystemic manifestations, and variability in its presentation. In 90-
95% of the cases reported, a gene deletion occurs at the 7q11.23 location on the long
arm of chromosome 7; in the remaining cases, a variety of other chromosomal
deletions and translocations have been observed (Wang et al., 1999). The most severe
cases are characterized by cardiac anomalies, including aortic stenosis, mental
retardation, growth deficiency, a characteristic facial appearance, dental
malformation, and infantile hypercalcemia (Lashkari et al., 1999).

[0149] The underlying molecular basis for the syndrome is the absence of
the proteins encoded by the genes of the affected region of the chromosome. A
missing elastin gene, with resulting extracellular matrix anomalies, is a consistent

finding. Other genes that are present in and near the commonly deleted region of chromosome 7, and thus are likely to contribute to pathogenesis, are (1) a gene encoding a regulator of chromosome condensation-like G-exchanging factor, which is a factor that exchanges nucleotides for small GTP-binding proteins, (2) an N-acetylgalactosaminyltransferase, (3) a DNAJ-like chaperone, (4) NOL1/NOP2/sun domain-containing proteins, including a novel protein designated WBSCR20, which is expressed in skeletal muscle, and is similar to a 120 kilodalton proliferation-associated nucleolar antigen, (5) a methyltransferase designated WBSCR22, and (6) other proteins with no known homologies (Merla et al., 2002; Doll and Grzeschik, 2001). Williams-Beuren-related sequences can possess or interact with a GTF2I-like repeat (GTF2I) domain, which is a DNA binding domain commonly deleted in Williams-Beuren syndrome, (http://pfam.wustl.edu/cgi-bin/getdesc?name=GTF2I).

Rheumatic Diseases

[0150] Rheumatic diseases are inflammatory conditions that can have autoimmune, infective, or traumatic origins. They include arthritis, systemic lupus erythematosus, scleroderma, and Sjogren's syndrome. Arthritis refers to any inflammation of a joint. Systemic lupus erythematosus is an autoimmune disease in which patients produce antibodies to their own tissues, resulting in an inflammatory process that can damage organs. Scleroderma can present as systemic scleroderma, a chronic, progressive disease that is characterized by hardening and stiffening of the skin and damage to internal organs, e.g., heart, lungs, kidneys and esophagus. Sjogren's syndrome is a progressive immunological disorder characterized by inflammation and the subsequent destruction of exocrine glands, e.g., salivary glands, sweat glands, and lacrimal (tear) glands.

[0151] The serum of patients with scleroderma and Sjogren's syndrome have antibodies directed against a protein that is a normal component of the Golgi apparatus (Seelig et al., 1994), an intracellular organelle composed of a stack of flattened cisternae with associated transport vesicles. The Golgi apparatus sorts proteins and sends them to their correct intracellular destination. This antigenic protein is a "golgin," one of a class of molecules characterized by an integral membrane domain and a large cytoplasmic region. Golgins organize the Golgi's structure, and influence protein sorting (Gillingham et al., 2002). Golgins function in a variety of ways, including cross-bridging Golgi cisternae to one another (Linstedt and Hauri, 1993) and tethering Golgi transport vesicles to the cisternal membranes

(Shorter et al., 2002). Rheumatic disease-associated sequences can possess or interact with golgin-97, RanBP2alpha, Imh1p, and p230/golgin (GRIP) domains, which are found in many large coiled-coil proteins, are sufficient for targeting to the Golgi, and have a conserved tyrosine residue (http://pfam.wustl.edu/cgi-bin/getdesc? name=GRIP).

Disintegrin-Related Sequences

[0152]   Disintegrins are proteins that interfere with the function of integrins. Disintegrins are generally proteins of about 70 amino acid residues that contain multiple disulfide bonds, bind with high affinity to a subset of integrins, and interfere with integrin binding to physiological ligands. Examples of disintegrin-related sequences include snake venoms and related proteins, cysteine-rich metalloproteinases and related non-enzymatic sequences, e.g., those expressed in the male reproductive tract, and membrane-anchored metalloproteinases with diverse functions, e.g., the shedding of cell-surface proteins such as cytokines and cytokine receptors, and the conferring of asthma susceptibility (Van Eerdewegh et al., 2002; Perry et al., 1995).

[0153]   Disintegrin-related sequences can possess or interact with disintegrin domains, which contain an Arg-Gly-Asp sequence, a sequence commonly found in adhesion proteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=disintegrin). Proteins that comprise both disintegrin and metalloproteinase peptidase domains include ADAM proteins. Disintegrin-related sequences can also possess or interact with reprolysin family propeptide (Pep_M12B_propep) domains, which are domains that include the propeptide sequence of members of the peptidase family M12B, and contain a sequence motif similar to a sequence found in matrixin proteins (http://pfam.wustl.edu/ cgi-bin/getdesc?name=Pep_M12B_ propep).

Factor-Related Sequences

[0154]   A factor is any molecule that contributes to a bodily process. Factors can function in specific biochemical reactions and cellular functions. There are many categories of factors, and factors are involved in many, if not all, physiological and pathological processes. Some exemplary factors are described in the following paragraphs; they are not exhaustive of the category.

[0155]   Transcription factors are factors that initiate or regulate transcription in eukaryotes. They include gene regulatory proteins, which turn specific sets of genes on or off, and general transcription factors, which assemble at the promoter

region to enable and regulate transcription of many genes. They also include transcription elongation factors, which are proteins required for the addition of amino acids to growing polypeptide chains on ribosomes (Alberts et al., 1994). Transcription factors interact with a wide variety of molecules, including DNA binding proteins, polymerases, regulatory molecules such as kinases, and specific regions of DNA, e.g., promoters, and enhancers (Alberts et al., 1994; Vallejo et al., 1993).

[0156]   Translation factors, including translation initiation factors and release factors, are involved in initiating and regulating the rate of protein synthesis. They also interact with many molecules, including ribosomal proteins, mRNA, and molecules that regulate the incorporation of amino acids into protein, such as kinases and GTP (Price et al., 1993; Alberts, 1994).

[0157]   Export factors are involved in the export of molecules, e.g., RNA, from the nucleus (Stutz et al., 2000). Folding factors are involved in the process of folding proteins into their functional three dimensional shapes, and are also involved in receptor function (Gao et al., 1994). Factors such as activators and coactivators interact with nuclear receptors to modulate cellular processes, e.g., transcription (Mahajan et al., 2002).

[0158]   ADP-ribosylation factors are involved in the addition of an ADP-ribose group donated from nicotinamide adenine dinucleotide (NAD) to specific amino acid residues in heterotrimeric G-proteins. They are involved in, for example, normal cellular processes, such as vesicular transport, and also in the pathologic states induced by cholera, pertussis, and botulinum toxins (Alberts et al., 1994; Amor et al., 1994). Guanine nucleotide exchange factors bind to small G-proteins, such as Ras, and displace GDP in favor of GTP. They act as effectors or modulators of small G-proteins (Ehrhardt et al., 2001; Janeway et al., 2001; Shao and Andres, 2000).

[0159]   Factor-related sequences can possess or interact with ADP-ribosylation factor family (arf) domains, which are GTP-binding domains involved in protein trafficking (http://pfam.wustl.edu/cgi-bin/getdesc?name=arf). Factor-related sequences can also possess or interact with elongation factor Tu GTP binding (GTP_EFTU) domains, which are elongation factors that promote the GTP-dependent binding of aminoacyl tRNA to ribosomes during protein biosynthesis, and catalyze the translocation of the newly synthesised protein chain (http://pfam.wustl.edu/cgi-bin/getdesc?name=GTP_EFTU). Factor-related sequences can also possess or

48

interact with 4F5 protein family (4F5) domains, which comprise ubiquitously expressed short proteins rich in aspartate, glutamate, lysine and arginine (http://pfam.wustl.edu/cgi-bin/getdesc?name=4F5). Factor-related sequences can also possess or interact with eukaryotic initiation factors, e.g., eukaryotic initiation factor 4E (IF4E), which recognizes and binds mRNA during an early step of protein synthesis (http://pfam.wustl.edu/cgi-bin/getdesc?name=IF4E).

Germ Cell Specific Protein-Related Sequences

[0160]   Germ cells, also called gametes, are cells that contribute to a new generation of organisms by giving rise to either an egg or a sperm. They are haploid cells specialized for sexual fusion. Proteins that are specific to germ cells can be found at one or more developmental stages of gametes.

[0161]   Germ cell-related sequences include germ cell genes and their gene products, their regulators and effectors, genes and gene products affected in disorders associated with germ cells, and antibodies that specifically recognize or modulate germ cell-related sequences. Examples of germ cell-related sequences include the germ cell-specific Y-box binding protein and contrin. Germ cell specific protein-related sequences possess or interact with the cold-shock DNA-binding (CSD) domain, which is described above.

Growth Factor-Related Sequences

[0162]   A growth factor is an extracellular polypeptide signaling molecule that stimulates a cell to grow or proliferate. Many types of growth factors exist, including protein hormones and steroid hormones. Some growth factors have a broad specificity, and some have a narrow specificity. Examples of growth factors with broad specificity include platelet-derived growth factor, epidermal growth factor, insulin like growth factor I, transforming growth factor β, and fibroblast growth factor, which act on many classes of cells. Examples of growth factors with narrow specificity include erythropoeitin, which induces proliferation of precursors of red blood cells, interleukin-2, which stimulates proliferation of activated T-lymphocytes, interleukin-3, which stimulates proliferation and survival of various types of blood cell precursors, and nerve growth factor, which promotes the survival and the outgrowth of nerve processes from specific classes of neurons.

[0163]   Most growth factors have other actions in addition to inducing cell growth or proliferation, e.g., they may influence survival, differentiation, migration,

49

or other cellular functions. Growth factors can have complex effects on their targets, e.g., they may act on some cells to stimulate cell division, and on others to inhibit it. They may stimulate growth at one concentration, and inhibit it an another. Growth factors are also involved in tumorogenesis.

[0164] Growth factor related sequences include sequences associated with the process of stimulating cell growth or proliferation by a growth factor. For example, they include intracellular effectors of growth, such as components of intracellular pathways that respond to growth factors (Kothapalli et al., 1997; Wax et al., 1994), sequences that bind directly or indirectly to growth factors (Van den Berghe et al., 2000), and sequences affected as a result of growth factor action.

[0165] Growth factor-related sequences can possess or interact with a transforming growth factor beta like (TGF-beta) domain, which is a multifunctional peptide sequence that controls proliferation, differentiation and other functions in many cell types (http://pfam. wustl.edu/cgi-bin/getdesc?name=TGF-beta). Growth factor-related sequences can also possess or interact with a fibroblast growth factor (FGF) domain, which is found in a family of proteins involved in growth and differentiation (http://pfam.wustl.edu/cgi-bin/getdesc? name=FGF).

GTPase-Related Sequences

[0166] GTPases are enzymes that catalyze GTP hydrolysis, and comprise a large family of proteins with a similar globular GTP binding domain. When GTP is bound to a GTPase, it is hydrolyzed to GDP, and the domain undergoes a conformational change that inactivates the protein. GTPases are regulated by GTPase regulators, proteins that determine whether a GTP binding protein exists in a GTP-bound or GDP-bound state (Zhao et al., 2003). GTPase regulators include GTPase activating proteins, which bind the GTPase and induce it to hydrolyze its bound GTP to GDP; the GTPase remains in an inactive, GDP-bound state until it encounters a guanine nucleotide releasing protein, which binds to the GTPase and causes the release of the nucleotide. GTPases have a broad spectrum of intracellular functions, including intracellular vesicular transport. Examples of GTPase-related sequences include ras, GTPase-activating proteins, and guanine nucleotide releasing proteins.

[0167] GTPase-related sequences can possess or interact with GTPase activator protein for Ras-like GTPase (RasGAP) domains, which are protein domains of about 250 residues that accelerate the GTPase activity of ras

(http://pfam.wustl.edu/cgi-bin/getdesc?name=RasGAP). GTPase-related sequences can also possess or interact with putative GTPase activating protein for ARF (ArfGap) domains, which are protein domains with a zinc finger involved in intermolecular associations (http://pfam.wustl.edu/cgi-bin/getdesc?name=ArfGap). GTPase-related sequences can also possess or interact with ankyrin repeat domains (ank), which are tandemly repeated modules of about 33 amino acids found in a variety of functionally diverse proteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=ank). GTPase-related sequences can also possess or interact with pleckstrin homology (PH) domains, which are protein domains of about 100 residues involved in intracellular signaling, or as components of the cytoskeleton (http://pfam.wustl.edu/cgi-bin/getdesc?name=PH).

Heat-Shock Protein-Related Sequences

[0168] Heat-shock proteins, also referred to as stress-response proteins, are proteins that are synthesized in response to an elevated temperature or other cell stressor, and help the cell withstand environmental insults. A cell stressor can induce a battery of genes that encode gene products that protect the cell from the result of the insult, e.g., proteins that stabilize and repair partially denatured cell proteins. Some heat-shock proteins, e.g., chaperones, are present at high levels in unstressed cells, and further induced by stress. Chaperones assist other proteins in attaining their proper secondary and tertiary structures. For example, members of the tubulin-specific chaperone A family possess tubulin-specific chaperone A (TBCA) domains that fold tubulin polypeptides into their functional configuration (http://pfam.wustl.edu/cgi-bin/getdesc?name=TBCA).

[0169] Heat and other stressors further induce the synthesis of a family of 90-kDa heat-shock proteins that are already abundant in unstressed cells (Pepin et al., 2001;Lees-Miller et al., 1989; Rebbe et al., 1987). Members of this family possess a hsp 90 protein (HSP90) domain that interacts with tubulin, actin, tyrosine kinase oncogene products of retroviruses, eIF2alpha kinase, and steroid hormone receptors (Lees-Miller and Anderson, 1989). This domain includes a highly-conserved N-terminal region, separated from a conserved, acidic C-terminal region by a highly-acidic, flexible linker region (http://pfam. wustl.edu/cgi-bin/getdesc?name=HSP90).

[0170] Another family of heat-shock proteins, the hsp70 proteins, have an average molecular weight of 70 kDa; some members of this family are only expressed under conditions of stress, while some are present in cells under normal conditions. Hsp70 proteins reside in different cellular compartments, e.g., the nucleus, cytosol,

mitochondria, and endoplasmic reticulum. Hsp70 proteins, e.g., Hsc73, can be differentially expressed at different stages of development (Soulier et al., 1996). Hsp70 proteins, e.g., the chaperone hsp70-like dnaK protein, can associate with proteins that possess a DnaJ domain, which comprises an N-terminal conserved domain of about 70 amino acids, a glycine-rich region of about 30 amino acids, a central domain containing four repeats of a CXXCXGXG motif, and a C-terminal region of 120 to 170 amino acids (http://pfam.wustl.edu/cgi-bin/getdesc? name=DnaJ). Proteins with DnaJ domains can be postranslationally modified by farnesylation (Andres et al., 1997).

Helicase-Related Sequences

[0171]   Helicases are enzymes that use energy from the hydrolysis of ATP to unwind the DNA helix at the replication fork, allowing the single stands to be copied. Proteins with DNA helicase activity play roles in DNA replication, repair, and recombination. Disorders associated with helicases include Xeroderma pigmentosum, Cockayne syndrome, diffuse collagen disease, alpha-thalassemia, Bloom syndrome, Werner syndrome, and Rothmund-Thomson syndrome (Miyajima, 2002). Examples of helicases include RNA helicases, RECQL4, and minichromosome maintenance helicase.

[0172]   Helicase-related sequences can possess or interact with helicase associated (HA) domains, which are protein domains comprising alpha helices that may bind to nucleic acids (http://pfam.wustl.edu/cgi-bin/getdesc?name=HA). Helicase-related sequences can also possess or interact with helicase conserved C-terminal (helicase_C) domains, which are protein domains that are found in a subset of helicases designated the DEAD/H helicases (http://pfam.wustl.edu/ cgi-bin/getdesc?name=helicase_C).

Hydrolase-Related Sequences

[0173]   Hydrolases are enzymes that catalyze the hydrolysis of a variety of bonds, such as esters, glycosides, and peptides. Hydrolases split a molecule into fragments by adding water; the water's hydrogen atom is incorporated into one fragment, and the hydroxyl group is incorporated into another. Hydrolases are involved in a wide range of physiological and pathological processes, including proteolysis, phosphatase activity, and sugar metabolism. Examples of hydrolases include protein hydrolases, lipid hydrolases, nucleic acid hydrolases, and small molecule, e.g., coenzyme A, hydrolases (Hawes et al., 1996).

[0174]   Hydrolase-related sequences can possess or interact with alpha/beta hydrolase fold (abhydrolase) domains, which are catalytic domains found in a wide range of hydrolytic enzymes of different phylogenetic origins and catalytic functions (http://pfam.wustl.edu/cgi-bin/getdesc?name=abhydrolase). Hydrolase-related sequences can also possess or interact with dUTPase domains, which are proteins domains that hydrolyze dUTP to dUMP and pyrophosphate.

Immune Cell-Related Sequences

[0175]   An immune cell is a cell involved in, or associated with, the immune system. Immune cells include cells in the myeloid and lymphocytic arms of the immune response, as well as their precursors. Immune cells also include cells at all stages in the differentiation pathways that produce cells associated with the immune system. These cells can reside, either permanently or temporarily, in the spleen, lymph nodes or mucosal-associated lymphoid tissues (MALT). Immune cell-related sequences are involved in all functions of the immune response, e.g., antibody production and cell-mediated immunity, and can function at any point in time, ranging from the embryonic formation of the immune system, through the time of an immune challenge, to many decades later, e.g., when a B-cell memory response is invoked (Janeway, 2001).

[0176]   Immune-cell related sequences of differentiating immune cells include pre-B cells that do not produce immunoglobulin light chain, but express a transcript homologous to immunoglobulin lambda light-chain genes, the expression of which is limited to pre-B cells and select other cells that have no surface immunoglobulin (Hollis et al., 1989). Immune-cell related sequences of activated immune cells include a B-cell-restricted transcription factor expressed by activated B cells; its expression pattern suggests it has a role in regulating B-cell differentiation (Massari et al., 1998).

[0177]   Examination of the expression of immune-cell related sequences can detect and diagnose immunoregulatory abnormalities. For example, genes that encode proteins which mediate the combinatorial process that combines a finite number of component genes into the very broad range of antigen-specific immunoglobulin and T-cell binding proteins, are expressed at higher levels in patients with systemic lupus erythematosis (SLE) than in healthy subjects (Girschick et al., 2002).

[0178] Immune cell-related sequences can possess or interact with a CUB domain, which is an extracellular domain of approximately 110 amino acids, and is present in functionally diverse, including developmentally regulated, proteins (http://pfam.wustl.edu/ cgi-bin/getdesc?name=CUB). Immune cell-related sequences can also possess or interact with a CD-20 domain, which has four transmembrane regions, both extracellular and cytoplasmic extensions, and is found, inter alia, in a high affinity IgE receptor (http://pfam.wustl.edu/cgi-bin/getdesc?name=CD20). Immune cell-related sequences can also possess or interact with an interferon-induced transmembrane protein (CD225) domain, which is found in a family of proteins that includes the human leukocyte antigen CD225, an interferon-inducible transmembrane protein associated with interferon-induced cell growth suppression (http://pfam.wustl. edu/cgi-bin/getdesc?name=CD225). Immune cell-related sequences can also possess or interact with sushi domains, also known as complement control protein (CCP) modules, or short consensus repeats (SCR). These domains are found in a wide variety of complement and adhesion proteins, including proteins responsible for the antigenicity of blood group antigens on the external face of the red blood cell membrane (http://pfam.wustl.edu/cgi-bin/getdesc?name=sushi). Immune cell-related sequences can also possess or interact with SH2 domains and rvt domains; both are described above.

Integrase-Related Sequences

[0179] Integrases are enzymes that form proviruses by inserting a linear double-stranded DNA copy of a retroviral genome into host cell DNA. Examples of integrases include HIV integrase, PhiC31 integrase, and Sip.

[0180] Integrase-related sequences can possess or interact with an integrase zinc binding domain (Integrase_Zn) domain, which is a zinc binding protein domain placed near the N-terminus (http://pfam.wustl.edu/cgi-bin/getdesc? name=Integrase_Zn). Integrase-related sequences can also possess or interact with an integrase core (rve) domain, which is a protein domain that forms the central catalytic core of the integrase (http://pfam.wustl.edu/ cgi-bin/getdesc?name=rve). This domain acts as an endonuclease to cleave the nucleotide and catalyzes the transfer of the viral DNA strand to the integration site of the host DNA. Integrase-related sequences also possess or interact with an integrase DNA binding (integrase) domain, which is a DNA-binding protein domain near the C-terminus (http://pfam.wustl.edu/cgi-bin/getdesc?name=integrase). Integrase-related sequences also possess or interact

reverse transcriptase (rvt) domains, which are described above. Integrase-related sequences also possess or interact with a RNase H domain, which is a protein domain that hydrolyzes the RNA portion of RNA/DNA hybrids (http://pfam.wustl.edu/cgi-bin/getdesc?name=rnaseH).

Integrin-Related Sequences

[0181]   Integrins are transmembrane proteins that mediate cell to cell as well as cell to matrix adhesion, and provide a means of communication between the interior of a cell and the extracellular matrix. The extracellular portion of integrins binds to components of the extracellular matrix, e.g., collagen, fibronectin and laminin. The intracellular portion of integrins interacts with the cell cytoskeleton, e.g., actin filaments near the cell surface. Integrins transmit information about the extracellular environment across the plasma membrane to the cytoskeleton, where it is available to intracellular signaling mechanisms (Alberts et al., 1994). Structurally, integrins consist of heterodimers of an alpha and a beta subunit. Each subunit has a large N-terminal extracellular domain followed by a transmembrane domain and a short C-terminal cytoplasmic region. The pairing of certain alpha subunits with certain beta-subunits determines ligand specificity, localization and function. The extracellular binding domains of integrins often bind their ligands with low affinity; simultaneous, weak, binding with multiple matrix molecules provides the cell with a means to sense its complex, changing, extracellular environment without becoming glued to it. Examples of integrin-related sequences include integrin alpha and beta subunits, collagens, and integrin-linked kinase (Zhang et al., 2002).

[0182]   Integrin-related sequences can possess or interact with von Willebrand factor type A (vwa) domains, which are protein domains that participate in diverse biological functions, e.g., cell adhesion, migration, homing, pattern formation, and signal transduction (http://pfam.wustl. edu/cgi-bin/getdesc?name=vwa). Integrin-related sequences can also possess or interact with FG-GAP repeat (FG-GAP) domains, which are protein domains present in the vicinity of ligand binding domains at the N-terminus of integrin alpha subunits (http://pfam.wustl.edu/cgi-bin/getdesc?name=FG-GAP).

Interacting Protein-Related Sequences

[0183]   An "interacting protein" is a protein that interacts with another molecule. Interacting proteins are involved in every aspect of cellular function. Interacting proteins have been characterized in all known locations in the cell, and

include all, or most types of, proteins. Interacting proteins in the nucleus regulate such diverse functions as apoptosis, transcription, homologous recombination, and DNA repair. Nuclear fibroblast growth factor-2 interacting factor interacts with fibroblast growth factor 2 to prevent apoptosis (Van den Berghe et al., 2000). Grap2 cyclin-D interacting protein (GCIP) a nuclear cell-cycle protein, inhibits select transcriptional events, and reduces the leve l of phosphorylation of nuclear retinoblastoma protein (Chang et al., 2000). Pir 51, a human homologue of Rec A, a bacterial enzyme that mediates genetic recombination, interacts with the enzyme rad51 to regulate homologous recombination and DNA repair in mammalian cells (Kovalenko et al., 1997). Hepatitis B virus X-associated protein (HBXAP), a protein demonstrated to play a role in the development of hepatocelluar carcinoma, interacts with the hepatitis B virus regulatory gene product HBx to increase viral transcription (Shamay et al., 2002).

[0184] Interacting protein-related proteins can utilize many protein domain motifs for interaction. They can possess or interact with domains that mediate interaction with DNA, RNA, ions, or other proteins. For example, PDZ domains, which are also known as DHR or GLGF domains, target signaling molecules to membranes and mediate the assembly of functional membrane domains (Fanning and Anderson, 1999). Interacting protein-related proteins can also possess or interact with rrm domains, which are described above.

Isomerase-Related Sequences

[0185] Isomerases are enzymes that convert molecules into their positional isomers, i.e., into molecules with the same chemical formula but a different stereochemical arrangement of atoms. Isomerases act on a wide variety of molecules, including sugars, amino acids, and nucleic acids. They are involved in a wide range of physiological and pathological functions, including those involving metabolic and synthetic pathways.

[0186] Isomerase-related sequences include isomerase genes and gene products, their substrates, products, activators, inhibitors, effectors, and cofactors, regulatory molecules that modulate their function, genes and gene products affected in disorders associated with isomerases and antibodies that specifically recognize or modulate isomerase-related sequences. Examples of isomerase-related sequences include triosephosphate isomerases, peptidyl-prolyl isomerases, glucose phosphate

isomerases, disulfide isomerases, ketosteroid isomerases, and ribosyltransferase-
isomerases (Brown et al., 1985).

[0187]   Isomerase-related sequences can possess or interact with
triosephosphate isomerase (TIM) domains, which are protein domains that catalyze
the reversible interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone
phosphate (http://pfam.wustl.edu/cgi-bin/getdesc?name=TIM). Isomerase-related
sequences can also possess or interact with cyclophilin type peptidyl-prolyl cis-trans
isomerase (pro_isomerase) domains, which accelerate protein folding by catalyzing
the cis-trans isomerization of peptide bonds (http://pfam.wustl.edu/
cgibin/getdesc?name=pro_ isomerase).

Mucin-Related Sequences

[0188]   The term mucin refers to both an albumin-like substance that is
present in mucus, and to transmembrane proteins that can typically be produced in
both soluble and transmembrane forms. Soluble mucins comprise mucus gels that
protect epithelial cells in the airways, digestive tract, and other organs, and are found
in body fluids, such as milk, tears, and saliva. In their transmembrane forms, mucins
provide a steric barrier to protect the apical surface of epithelial cells.
Transmembrane mucins are also involved in pathogenesis; for example, they mediate
viral entry into cells, promulgate the inflammatory response, and are involved in the
regulation of abnormal cell proliferation (Jeffery and Zhu, 2002; Tsuda et al., 1993).
Examples of mucins include MUC2 mucin, mucin carcinoembryonic antigen, and
Muc3 membrane bound intestinal mucin.

[0189]   Mucin -related sequences can possess or interact with mucin-like
glycoprotein (tryp_mucin) domains, which are domains that are involved in the
interaction of parasites with host cells (http://pfam.wustl.edu/cgi-
bin/getdesc?name=Tryp_mucin). Mucin-related sequences can also possess or
interact with multi-glycosylated core protein (MGC-24) domains, which are protein
domains of sialomucins that are expressed in many normal and cancerous tissues
(http://pfam.wustl.edu/cgi-bin/getdesc?name=MGC-24).

Other Polypeptide-Related Sequences

[0190]   In addition to the sequences described above, the sequences of
the invention include nucleotide and amino acid sequences, some with known
function, and some with unknown function, that fall into a broad array of categories.

These sequences are listed below in SEQ ID NOS.: 1232 - 2462, as "Other Polypeptides with Known Function," and "Other Polypeptides," respectively.

[0191] Polypeptide-related sequences of the invention can possess or interact with groucho/TLE N-terminal Q-rich (TLE_N) domains, which are protein domains found in co-repressor proteins, and are involved in oligomerization (http://pfam.wustl.edu/cgi-bin/getdesc?name=TLE_N). Polypeptide-related sequences of the invention can also possess or interact with uncharacterized protein family 0160 (UPF0160) domains, which are protein domains found in proteins that include multiple metal-binding residues, and in some cases act as a phosphodiesterase (http://pfam.wustl.edu/cgi-bin/getdesc?name=UPF0160). Polypeptide-related sequences of the invention can also possess or interact with SNF7 domains, which are protein domains involved in protein sorting and transport from the endosome to the lysosome or vacuole of eucaryotic cells (http://pfam.wustl.edu/cgi-bin/getdesc?name=SNF7). Polypeptide-related sequences of the invention can also possess or interact with NifU-like N-terminal (NifU_N) domains, which are protein domains involved in nitrogen fixation, and other functions (http://pfam.wustl.edu/cgi-bin/getdesc? name=NifU_N). Polypeptide-related sequences of the invention can also possess or interact with tRNA synthetases class II (D, K, and N) (tRNA-synt_2) domains, which are protein domains that activate the amino acids asparagines, aspartic acid, and lysine, and transfer them to specific tRNA molecules (http://pfam.wustl.edu/cgi-bin/getdesc?name=tRNA-synt_2).

[0192] Polypeptide-related sequences of the invention can also possess or interact with dynein heavy chain (dynein_heavy) domains, which are protein domains that correspond to the C-terminal region of the dynein heavy chain (http://pfam.wustl.edu/cgi-bin/getdesc?name=Dynein_heavy). Polypeptide-related sequences of the invention can also possess or interact with cyclin-dependent kinase regulatory subunit (CKS) domains, which are protein domains of approximately 79-150 amino acid residues that are involved in regulating progression through the cell cycle (http://pfam.wustl.edu/cgi-bin/getdesc?name= CKS).

[0193] Polypeptide-related sequences of the invention can also possess or interact with nucleoside diphosphate linked to some other moiety X (NUDIX) domains, which are protein domains that are involved in removing oxidatively damaged nucleotides (http://pfam.wustl.edu/cgi-bin/getdesc?name=NUDIX). Polypeptide-related sequences of the invention can also possess or interact with T-

complex protein/cpn60 chaperonin (cpn60_TCP1) domains, which are protein domains involved in protein folding and oligomerization (http://pfam.wustl.edu/cgi-bin/getdesc?name=cpn60_TCP1). Polypeptide-related sequences of the invention can also possess or interact with F-actin capping protein, beta subunit (F_actin_cap_B) domains, which are protein domains of approximately 280 amino acids that are involved in capping actin, i.e., blocking the exchange of actin monomers (http://pfam. wustl.edu/cgi-bin/getdesc?name=F_actin_cap_B).

[0194]   Polypeptide-related sequences of the invention can also possess or interact with G-protein alpha subunit (G-alpha) domains, which are protein domains that bind guanyl nucleotides, and function as a GTPase (http://pfam.wustl. edu/cgi-bin/getdesc? name=G-alpha). Polypeptide-related sequences of the invention can also possess or interact with Kruppel-associated box (KRAB) domains, which are protein domains involved in protein-protein interactions, and present in some zinc finger proteins (http://pfam.wustl.edu/ cgi-bin/getdesc?name=KRAB). Polypeptide-related sequences of the invention can also possess or interact with metallopeptidase family M24 (Peptidase_M24) domains, which are protein domains that are found in some metalloproteases, including proline dipeptidase, and methionine aminopeptidase (http://pfam.wustl.edu/cgi-bin/getdesc?name=Peptidase_M24). Polypeptide-related sequences of the invention can also possess or interact with thioredoxin (thiored) domains, which are protein domains involved in oxidation/reduction reactions by reversibly oxidizing disulfide bonds (http://pfam.wustl.edu/cgi-bin/getdesc? name=thiored).

[0195]   Polypeptide-related sequences of the invention can also possess or interact with TUDOR domains, which are protein domains involved in the formation of primordial germ cells, and for normal abdominal segmentation (http://pfam.wustl.edu/cgi-bin/getdesc?name =TUDOR). Polypeptide-related sequences of the invention can also possess or interact with SIT4 phosphatase-associated protein (SAPS) domains, which are protein domains that are involved in cyclin transcription (http://pfam.wustl.edu/cgi-bin/getdesc?name=SAPS). Polypeptide-related sequences of the invention can also possess or interact with ankyrin repeat (ank) domains, which are protein domains of approximately 33 amino acids, and are sometimes found in tandemly repeated modules (http://pfam.wustl.edu/ cgi-bin/getdesc? name=ank). Polypeptide-related sequences of the invention can also possess or interact with nicotinamide N-methyltransferase/phenylethanolamine N-

methyltransferase/ thioether S-methyltransferase (NNMT_PNMT_TEMT) domains, which are protein domains that are found in proteins that use S-adenosyl-L-methionine as the methyl donor (http://pfam.wustl.edu/cgi-bin/getdesc?name= NNMT_PNMT_TEMT). Polypeptide-related sequences of the invention can also possess or interact with C1q domains, which are protein domains involved in activating the serum complement system (http://pfam.wustl.edu/cgi-bin/getdesc? name=C1q). Polypeptide-related sequences of the invention can also possess or interact with collagen triple helix repeat (Collagen) domains, which are protein domains that typically form extracellular connective tissue (http://pfam.wustl.edu/cgi-bin/getdesc? name=Collagen).

[0196] Polypeptide-related sequences of the invention can also possess or interact with the hyaluronan/mRNA binding family (HABP4_PAI-RBP1) domain, which is a protein domain that can bind to the glucosaminoglycan hyaluronan, and to RNA (http://pfam.wustl.edu/cgi-bin/getdesc?name=HABP4_PAI-RBP1). Polypeptide-related sequences of the invention can also possess or interact with eucaryotic aspartyl protease (asp) domains, which are protein domains that cleave peptide bonds; proteins with this domain include pepsins, cathepsins, and rennin (http://pfam.wustl.edu/cgi-bin/getdesc?name=asp). Polypeptide-related sequences of the invention can also possess or interact with trypsin domains, which are protein domains that function as serine proteases (http://pfam.wustl.edu/ cgi-bin/getdesc? name=trypsin). Polypeptide-related sequences of the invention can also possess or interact with Kunitz/Bovine pancreatic trypsin inhibitor (Kunitz_BPTI) domains, which are protein domains that is found in serine protease inhibitors (http://pfam. wustl.edu/cgi-bin/getdesc?name=Kunitz_BPTI). Polypeptide-related sequences of the invention can also possess or interact with proliferating cell nuclear antigen, N-terminal (PCNA) domains, which are protein domains that are found on non-histone acidic nuclear proteins, and play a role in controlling DNA replication (http://pfam. wustl.edu/cgi-bin/getdesc?name=PCNA).

Oxygenase-Related Sequences

[0197] Oxygenases are enzymes that catalyze the incorporation of molecular oxygen into organic substances. Dioxygenases, also known as oxygen transferases, catalyze the introduction of both atoms of molecular oxygen, and typically contain iron. Monooxygenases, also known as mixed function oxygenases, introduce one oxygen atom; the other is reduced to water. Examples of oxygenase-

related sequences include cytochrome oxygenases, heme oxygenases, cyclooxygenases, lipoxygenases, and peptide-aspartate beta-dioxygenase.

[0198]   Oxygenase-related sequences can possess or interact with alkyl hydroperoxide reductase/thiol specific antioxidant (AhpC-TSA) domains, which are responsible for providing a defense against sulfur-containing radicals; proteins that possess this domain include allergens, e.g., asp f 3, mal f 2, and mal f 3 (http://pfam.wustl.edu/cgi-bin/getdesc?name=AhpC-TSA). Oxygenase-related sequences can also possess or interact with monooxygenase domains, which are protein domains that utilize flavin adenine dinucleotide (FAD) (http://pfam.wustl.edu/cgi-bin/getdesc?name=Monooxygenase). Oxygenase-related sequences can also possess or interact with dioxygenase domains, which are protein domains that catalyze the incorporation of both atoms of molecular oxygen into substrates (http://pfam.wustl.edu/cgi-bin/getdesc?name= Dioxygenase).

Peroxidase-Related Sequences

[0199]   Peroxidases are enzymes that catalyze the reduction of hydrogen peroxide. Peroxidases are generally located within peroxisomes, which are intracellular organelles that metabolize fatty acids and toxic compounds. Disorders associated with peroxidase-related sequences include X-linked adrenoleukodystrophy. Examples of peroxidase-related sequences include glutathione peroxidases, thiol peroxidases, catalases, horseradish peroxidases, anionic peroxidases, and thyroid peroxidases.

[0200]   Peroxidase-related sequences can possess or interact with alkyl hydroperoxide reductase/thiol specific antioxidant (AhpC-TSA) domains, which are protein domains that can reduce organic hydroperoxides (http://pfam.wustl.edu/cgi-bin/getdesc? name=AhpC-TSA).

Phospholipase-Related Sequences

[0201]   Phospholipases are enzymes that act on phospholipids. They characteristically generate products that are active in signal transduction pathways. For example, phospholipase C hydrolyzes phosphatidylinositol bisphosphate (PIP$_2$) to generate the two intracellular mediators, inositol trisphosphate (IP$_3$) and diacylglycerol. IP$_3$ releases Ca^{2+} from stores in the endoplasmic reticulum, increasing the cytosolic Ca^{2+} concentration. Diacylglycerol remains in the plasma membrane and activates protein kinase C.

[0202] Phospholipase activity is involved in the synthesis of eicosanoids, inflammatory mediators that include prostaglandins, prostacyclins, thromboxanes, and leukotrienes. Corticosteroid hormones, such as cortisone, for example, inhibit phospholipase activity in the first step of the eicosanoid synthesis pathway. Corticosteroid hormones are widely used clinically to treat noninfectious inflammatory diseases, such as some forms of arthritis (Ribardo et al., 2002).

[0203] Phospholipids play a pivotal role in the modulation of intestinal inflammation. The mucosal surface of the digestive tract functions as a regulatory barrier between the gastrointestinal lumen and the underlying mucosal immune system. Phospholipids help preserve the mucosa following various forms of injury or physiological damage to the lumen, thus preventing invasion of harmful luminal factors into the host, which subsequently may lead to inflammation, or a pathological immune response, both promoting and inhibiting gastrointestinal inflammation and immunity (Sturm and Dignass, 2002).

[0204] Phospholipase-related sequences can possess or interact with lysophospholipase catalytic (PLA2_B) domains, which catalyze the release of fatty cids from lysophospholipids (http://pfam.wustl.edu/cgi-bin/getdesc?name=PLA2_B). Phospholipase-related sequences can also possess or interact with phospholipase/carboxylesterase (abhydrolase_2) domains, which have broad substrate specificity (http://pfam.wustl.edu/cgi-bin/getdesc?name=abhydrolase_2). Phospholipase-related sequences can also possess or interact with GDSL-like lipase/acylhydrolase (Lipase_GDSL) domains, which are present in lipolytic enzymes with serine in the active site (http://pfam.wustl.edu/cgi-bin/getdesc?name= Lipase_GDSL).

Prosaposin-Related Sequences

[0205] Saposins are small lysosomal proteins that activate lysosomal lipid-degrading enzymes, including enzymes that metabolize sphingosine. They typically isolate lipids from their membrane surroundings, and increase their accessibility to degradative enzymes. Mammalian saposins are synthesized as a single precursor molecule, prosaposin, which becomes an active saposin following proteolytic activation. Examples of prosaposin-related sequences include saposin A, saposin B, and saposin C. Disorders associated with prosaposin-related sequences include neurodegenerative diseases similar to similar to Tay-Sachs and Sandhoff diseases, e.g., Gaucher's disease, which is described above.

[0206]   Prosaposin-related sequences can possess or interact with saposin-A (SAPA) domains, saposin B1 (SapB_1) domains, and saposin B2 (SapB_2) domains, which are described above.

Proteasome-Related Sequences

[0207]   Proteasomes are intracellular complexes that degrade proteins. Proteasomes recognize proteins that have been marked for destruction by the addition of an ubiquitin molecule, unfold these ubiquitinated proteins, cleave them into small peptides of 6-12 amino acids, and release them into the cytosol (Mitch and Goldberg, 1996). Examples of proteasome-related sequences include 26S proteasome subunits, 26S proteasome regulatory chains, and ubiquitin.

[0208]   Proteasome-related sequences can possess or interact with proteasome/cyclosome repeat (PC_rep) domains, which are protein domains that are present in regulatory subunits of the proteasome (http://pfam.wustl.edu/cgi-bin/getdesc?name= PC_rep). Proteasome-related sequences can also possess or interact with Mov34/MPN/PAD-1 family (Mov34) domains, which are protein domains found at the N-terminus of regulatory subunits of the proteasome (http://pfam.wustl.edu/cgi-bin/getdesc?name=Mov34).

Reductase-Related Sequences

[0209]   Reductases are enzymes that catalyze reduction reactions, i.e., reactions in which hydrogen is combined with a molecule, or reactions in which oxygen is removed from a molecule. Examples of reductases include dehydrogenase reductases, oxidoreductases, quinone reductases, CoA reductases, dihydrofolate reductases, tetrahydrofolate reductases, carbonyl reductases, nitrate reductases, epoxide reductases, NADP(+) reductases, ribonucleotide reductases, and thioredoxin reductases (Loeffen et al., 1998).

[0210]   Reductase-related sequences can possess or interact with short chain dehydrogenase (adh_short) domains, which are present in a wide variety of proteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=adh_short). Reductase-related sequences can possess or interact with NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus (oxidored_q1_N) domains, which are protein domains that catalyze the transfer of electrons from NADH to ubiquinone in a reaction that can be associated with proton translocation across a membrane (http://pfam.wustl.edu/cgi-bin/getdesc?name=oxidored_q1_N).

Reverse Transcriptase-Related Sequences

[0211]   Reverse transcriptases are enzymes that make double stranded DNA copies from single stranded nucleic acid template molecules. Typically, a reverse transcriptase is a DNA polymerase that can copy both RNA and DNA templates, and has an integral RNase H activity (Lim et al., 2002). The two enzymatic domains of reverse transcriptase reflect these two activities; the first is a DNA polymerase domain that can use either RNA or DNA as a template to synthesize either the minus-strand or the plus strand of DNA, and the second is an RNase H domain that degrades the RNA in RNA-DNA hybrids (Coffin, 1997; Wu and Gallo, 1975).

[0212]   Reverse transcriptase plays a role in the replication of some viruses, e.g., retroviruses. It copies the retroviral RNA genome to produce a single minus strand of DNA, then catalyzes the synthesis of a complementary plus strand. Accordingly, reverse transcriptase is a therapeutic target for conditions that involve retroviruses, e.g., Aquired Immune Deficiency Syndrome (AIDS). A number of anti-retroviral drugs inhibit reverse transcriptase (Frank, 2002).

[0213]   Reverse transcriptase is also a standard scientific research tool in the field of molecular biology. The reverse transcriptase polymerase chain reaction (RTPCR) amplifies specific DNA sequences rapidly, and in vitro. RTPCR can detect trace amounts of RNA and DNA, and is used in a wide range of applications, including forensics, the diagnosis of genetic diseases, determination of the prognosis of diagnosed diseases, and the detection of viral infection (Alberts, et al., 1994). For example, reverse transcriptase is used to diagnose cancer (Rowland, 2002), and to provide prognostic information about the predicted survival of patients with prostate cancer (Kantoff et al., 2001).

[0214]   An example of a reverse transcriptase is telomerase, a general tumor marker with a reverse transcriptase catalytic subunit (Kirkpatrick and Mokbel, 2001). Most human somatic cells do not express the telomerase reverse transcriptase gene; conversely, most cancer cells express this gene (Ducrest et al., 2002; Kyo et al., 2000). The human telomerase reverse transcriptase promoter has been placed in gene therapy vectors that specifically target telomerase-positive tumor cells, and spare nearby telomerase-negative cells (Pan and Koeneman, 1999). Human telomerase reverse transcriptase is also recognized as a tumor antigen that can be a target for immunotherapeutic approaches to cancer (Gordan and Vonderheide, 2002).

[0215]   Reverse transcriptase-related sequences can possess or interact with rvt, transposase_22, WD40, and Exo_endo_phos domains, all of which are described above.

Ribosome-Related Sequences

[0216]   A ribosome is a particle comprised of ribosomal proteins and ribosomal RNA that catalyzes protein synthesis from messenger RNA. Ribosomes are composed of two subunits, the large (L) subunit and the small (S) subunit. The typical mammalian ribosome comprises four RNA molecules and approximately eighty different proteins, which are highly conserved among prokaryotes and eukaryotes, and perform a variety of tasks related to protein synthesis . e.g., coordinating protein synthesis in a manner that maintains cell homeostasis (Yoshihama et al., 2002; Kenmochi et al., 1998).

[0217]   Ribosomal proteins can perform functions independent of their involvement in protein synthesis. For example, they are involved in cell-cycle progression, e.g., as cell cycle checkpoints, and mediators of homologous recombination, embryogenesis, and skeletal development (Yoshihama et al., 2002; Chen and Ioannou, 1999). They also contribute to the regulation of cell growth, transformation, and death, and can induce apoptosis (Chen and Ioannou, 1999; Naora et al., 1999). Mutations in ribosomal proteins are associated with human diseases, including Down syndrome, Diamond-Blackfan anemia, Turner syndrome, and Noonan syndrome (Yoshihama et al., 2002).

[0218]   Ribosomal proteins have been grouped into protein families on the basis of sequence similarities in functional domains. One family of ribosomal proteins, the ribosomal protein L11, RNA binding (Ribosomal_L11) domain, is comprised of members that possess the L11 RNA binding domain; this family includes the ribosomal proteins L11 and L12, which are components of the large subunit. L11 is a protein of 140 to 165 amino-acids that binds to a 23S RNA molecule, the C-terminal region of which is buried within the ribosomal structure (http://pfam.wustl.edu/cgi-bin/getdesc?name=Ribosomal_L11). Another family of large ribosomal subunit proteins possess the ribosomal protein L13e (Ribosomal_L13e) domain, which is found in a wide range of vertebrates and in lower-order species (http://pfam.wustl.edu/cgi-bin/getdesc?name=Ribosomal_L13e), as is the ribosomal protein L44 (Ribosomal_L44) domain (http://pfam.wustl.edu/cgi-bin/getdesc?name= Ribosomal_L44).

[0219]   Additional ribosomal protein families encompass small subunit proteins. The ribosomal protein S6e (Ribosomal_S6e) domain is present in a family of proteins which includes protein kinase substrates that control cell growth and proliferation by selectively translating particular classes of mRNA (http://pfam.wustl.edu/cgi-bin/getdesc?name= Ribosomal_S6e). The ribosomal protein S8e (Ribosomal_S8e) domain is present in a family of proteins comprising approximately 220 amino acids in eukaryotes, and about 125 amino acids in archebacteria (http://pfam.wustl.edu/cgi-bin/getdesc?name=Ribosomal_S8e). The ribosomal protein S10p/S20e (Ribosomal_S10) domain is present in a family of proteins which includes the small ribosomal subunit S10 from prokaryotes and S20 from eukaryotes (http://pfam.wustl.edu/cgi-bin/getdesc?name= Ribosomal_S10). S10 is involved in binding transfer RNA to the ribosome, and also operates as a transcriptional elongation factor.

RNase-Related Sequences

[0220]   RNases are enzymes that cleave RNA. RNases generally recognize their targets by tertiary structure, rather than by sequence; they include exonucleases, which remove the terminal base in an RNA sequence, and endonucleases, which can cleave non-terminal bases. Examples of RNases include RNase E, which is involved in the formation of 5S ribosomal RNA from pre-ribosomal RNA; RNase F, which cleaves both viral and host RNA in response to interferons, inhibiting protein synthesis; RNase H, which is specific for the RNA strand of an RNA-DNA hybrid; RNase P, which generates transfer RNA from precursor transcripts; and RNase T, which removes the terminal AMP from nonaminoacylated tRNA (Coffin, et al., 1997).

[0221]   RNase-related sequences can possess or interact with rvt, rve, RNase H, and gag_p30 domains, all of which are described above.

RNase H-Related Sequences

[0222]   RNase H is a nuclease specific for the RNA strand of an RNA-DNA hybrid that cleaves phosphodiester bonds to produce molecules with 3'-OH and 5'-PO_4 ends. Multiple forms of RNase H are present in both prokaryotes and eukaryotes. RNase H may be part of larger polypeptides and its activity can be influenced by other regions of these polypeptides (Coffin, et al., 1997; Crouch 1990).

[0223]   During retroviral replication, RNase H activity forms oligonucleotides that prime DNA synthesis. Therefore, the RNase H activity of

reverse transcriptase is a target for therapeutic intervention. For example, small molecule inhibitors of retroviral RNase H function have shown promise in managing HIV infection (Klarman, et al., 2002).

[0224] Another therapeutic indication for RNase H is the regulation of cancer genes by targeting mRNA translation. Antisense deoxyoligonucleotides down-regulate mRNA expression by annealing to specific regions of an mRNA. Formation of the DNA:RNA heteroduplex then triggers mRNA cleavage by RNase H. Cleavage is rapidly followed by further degredation, irreversibly preventing translation of the target mRNA. Antisense deoxyoligonucleotides that trigger RNase H activity can thus be used as cancer therapeutic agents (Crooke, 1996; Curcio et al., 1997).

[0225] RNase H-related sequences can possess or interact with rnaseH, Gag_p30, rvt, and rve domains, all of which are described above.

SH3-Related Sequences

[0226] Src homology region 3 (SH3) is a polypeptide domain commonly found in intracellular signaling proteins; it binds with moderate affinity and selectivity to proline-rich ligands. SH3 domains are heterogeneous; different SH3 domains bind to different proline-rich sequences (Gmeiner and Horita, 2001). SH3 domains are involved in a wide variety of biological processes, including mediating the assembly of large multiprotein complexes, regulating enzyme activity, and modulating the local concentration or subcellular localization of signaling pathway components (Mayer, 2001). Examples of SH3-related sequences include phosphotyrosine receptors, membrane associated guanylate kinases, mitogen-activated protein kinases, myosin 1, the Crk adaptor protein, phospholipase C-γ, Grb2, Sos, src-SH3, Abl-SH3, the Nck adaptor, and alpha-spectrin-SH3.

[0227] SH3-related sequences can possess or interact with SH3 domains, which are protein domains of approximately 50-70 amino acids, and are present in a large number of proteins involved in intracellular signaling (http://pfam. wustl.edu/cgi-bin/getdesc?name=SH3). SH3-related sequences can also possess or interact with SH3 domain-binding protein 5 (SH3BP5) domains, which are protein domains that act as a substrate for c-Jun N-terminal kinase (http://pfam.wustl.edu/cgi-bin/getdesc?name=SH3BP5).

Stem Cell-Related Sequences

[0228]   Stem cells are pluripotent or multipotent cells that generate maturing cells in multiple differentiation lineages. Pluripotent cells have the capacity to differentiate into each and every cell present in the organism. Embryonic stem cells are pluripotent; they can differentiate into any of the cells present in the adult. Multipotent cells have the ability to differentiate into more than one cell type. Organ-specific stem cells are multipotent; they can differentiate into any of the cells of the organ they inhabit.

[0229]   When they divide *in vivo*, both pluripotent and multipotent stem cells can maintain their pluripotency or multipotency while giving rise to differentiated progeny. Thus, stem cells can produce replicas of themselves which are pluri- or multipotent, and are also able to differentiate into lineage-restricted committed progenitor cells. For example, hematopoeitic stem cells, which are multipotent cells specifically able to form blood cells, can divide to produce replicate hematopoeitic stem cells. They can also divide to produce more highly differentiated cells, which are precursors of blood cells. The precursors differentiate, sometimes through several generations of cells, into blood cells. A hematopoetic stem cell can also divide into a cell with the capacity to form, for example, a relatively undifferentiated cell that is committed to differentiate into, i.e., granulocytes, or erythrocytes, or another type of blood cell.

[0230]   Stem cells can also reproduce and differentiate *in vitro*. Embryonic stem cells have been directed to differentiate into cardiac muscle cells *in vitro* and, alternatively, into early progenitors of neural stem cells, and then into mature neurons and glial cells in vitro (Trounson, 2002).

[0231]   Stem cell therapy is effective in treating cancer in humans (Slavin et al., 2001), and offers several advantages over traditional cancer therapies (Weissman, 2000). One advantage of stem cell therapy exists when used in conjunction with radiation therapy. In radiation therapy for cancer, the dose of radiation necessary to kill the cancer cells in an organ can also be sufficient to destroy the healthy cells of the organ. In combined stem cell and radiation therapy, an organ is first treated with sufficient radiation to destroy all of the cancer cells and most or all of the healthy cells, but then stem cells are infused to repopulate the organ. In the ensuing weeks, as the cancer cells and healthy cells die, the stem cells replace the healthy cells. Another advantage of this approach, compared to heterologous organ transplants, is that there

is no risk of rejection, since stem cells do not provoke an immune response. A further advantage is that stem cells are inherently programmed to regulate their numbers and differentiation status, i.e., once provided to the patient, the necessary number will differentiate, and the rest will remain undifferentiated (Weissman, 2000).

[0232] Stem cell therapy is also effective in treating autoimmune disease in humans. For example, immunosuppression in conjunction with stem-cell transplantation has induced remission in patients with refractory, severe rheumatic autoimmune disease (Van Laar and Tyndall, 2003). Patients with rheumatoid arthritis, systemic lupus erythematosus, systemic sclerosis, and juvenile idiopathic arthritis have benefited from stem cell transplants (Van Laar and Tyndall, 2003).

[0233] Preclinical studies also suggest the potential of stem cell transplantation for the treatment of neural and muscular injuries and disorders, including those of the central nervous system, peripheral nervous system, and skeletal, cardiac and smooth muscle (Deasy and Huard, 2002). Stem cells transplanted into the bone marrow of mice migrate to the site of injured muscle and differentiate into new muscle cells. For example, patients with myasthenia gravis, muscular dystrophies, amyotrophic lateral sclerosis, congestive heart failure, Parkinson's disease, and Alzheimer's disease may benefit from stem cell therapy (Henningson, 2003).

[0234] In addition to therapeutic uses, research using stem cells can provide useful information about normal stem cell function and the pathogenesis of disease. Stem cells derived from a patient with a genetic disease can provide a tool for studying that disease. To derive these stem cells, a somatic cell, i.e., a cell that is not in the oocyte or spermatocyte lineage, is donated by the patient, and the nucleus is removed and transferred to an unfertilized human oocyte. This nuclear transplant procedure produces, at the blastocyst stage of development, embryonic stem cells with the same set of genes as the patient with the genetic disease. Studying these cells, and their progeny in vitro, permits analysis of a specific model of the disease. For example, placing stem cells derived from a patient with a genetic disorder under the control of various stem cell regulatory factors can elicit abnormal responses from the affected stem cells compared to stem cells derived from a healthy individual's somatic nucleus.

[0235] Embryonic stem cell-related sequences can possess or interact with the stem cell factor (SCF) domain, a transmembrane domain having a soluble, secreted form, which is involved in hematopoeisis, and which binds to and activates a

receptor tyrosine kinase, stimulating the proliferation of mast cells and augmenting the proliferation of myeloid and lymphoid hematopoietic progenitors in bone marrow culture (http://pfam.wustl.edu/cgi-bin/getdesc?name=SCF).

[0236]   Certain stem cell related sequences can possess the ability to maintain the stem cell in undifferentiated state while allowing cell proliferation. Such compositions can be useful in *ex vivo* cell therapy to expand populations of cells for cell replacement therapy.

[0237]   Certain stem cell related sequences can possess the ability to cause cell differentiation to a relatively mature cell type and are useful to *in vivo* or *ex vivo* therapy to compensate for deficiency of such relatively mature cell type.

Synthetase-Related Sequences

[0238]   A synthetase is an enzyme that catalyzes the synthesis of a molecule. Synthetases comprise a broad class of enzymes; they catalyze the synthesis of nucleic acids, peptides, and lipids (Agou et al., 1996). Examples of synthetases include lysyl-tRNA synthetase, asparaginyl t-RNA synthetase, holocarboxylase synthetase, carbamyl phosphate synthetase I, and argininosuccinate synthetase.

[0239]   Synthetase-related sequences can possess or interact with transfer RNA synthetase domains, which are protein domains that activate amino acids and transfer them to specific transfer RNA molecules as a step in protein biosynthesis (http://pfam.wustl.edu/cgi-bin/getdesc?name=tRNA-synt_2). The 20 aminoacyl-tRNA synthetases are divided into class I and class II, each of which contain multiple synthetases with different specificities. For example, there is a protein domain involved in the asparagines, aspartic acid, and lysine synthesis (http://pfam.wustl. edu/cgi-bin/textsearch?terms=trna-synt&search_what=all& sections= DE&sections=CC&size=100). Synthetase-related sequences can also possess or interact with lipid-A-disaccharide synthetase (LpxB) domains, which are protein domains that catalyze the synthesis of disaccharides (http://pfam.wustl.edu/cgi-bin/getdesc? name=LpxB).

TATA Box-Related Sequences

[0240]   A TATA box is a consensus sequence in the promoter region of many eucaryotic genes that binds a general transcription factor and plays a role in specifying the position for transcription initiation. TATA boxes are generally found approximately 25 nucleotides before the site of transcription initiation (Chalut et al.,

70

1995). Examples of TATA box-related sequences include TATA box binding
protein, 13 TATA/TBP, and small nuclear RNA-activating protein 190 Myb DNA.

[0241]   TATA box-related sequences can possess or interact with
transcription factor TFIID, also known as the TATA-binding protein (TBP) domain,
which is a protein domain that specifically binds to the TATA box promoter element
(http://pfam.wustl.edu/cgi-bin/getdesc?name=TBP). TATA box-related sequences
can also possess or interact with HMG14 and HMG17 (HMG14_17) domains, which
are members of a family of high mobility group proteins, described above
(http://pfam.wustl. edu/cgi-bin/getdesc? name=HMG14_17).

Tat-Related Sequences

[0242]   Tat is a human immunodeficiency virus (HIV) protein involved
in viral production of new RNA genomes and new complete viral particles. Tat is
also involved in AIDS pathogenesis; it plays a role in reactivating latent viruses, e.g.,
the JC retrovirus; it is involved in the development of AIDS-related Kaposi's
Sarcoma; and it depresses the function of, and induces apoptosis in, helper CD4 cells
(Yu et al., 1995). Examples of Tat-related sequences include Tat-associated proteins,
e.g., Tap, HIV-1 Rev, and tat-associated kinase (also known as positive transcriptional
elongation factor b).

[0243]   Tat-related sequences can possess or interact with
transactivating regulatory protein (Tat) domains, which are protein domains that
contribute to efficient transcription of a viral genome (http://pfam.wustl.edu/cgi-
bin/getdesc?name=Tat). Tat-related sequences can also possess or interact with
mitochondrial glycoprotein (MAM33) domains, which are protein domains found in
mitochondrial matrix proteins, and which can be involved in mitochondrial oxidative
phosphorylation and in interactions between the nucleus and the mitochondria
(http://pfam.wustl.edu/cgi-bin/getdesc?name=MAM33).

Transferase-Related Sequences

[0244]   Transferases are enzymes that transfer a designated group of
atoms from a donor molecule to an acceptor molecule. For example, acyl transferases
transfer acyl groups, methyl transferases transfer methyl groups, nucleotidyl
transferases transfer nucleotides, prenyltransferases transfer prenyl groups, and
glycosyl transferases transfer glycosyl groups (Lin et al., 1996). Examples of
transferases include acetyltransferases, hydroxymethyltransferases, sialyltransferases,

arginine N-methyltransferase, glucoronosyltransferase, NTP-transferase, and GDP-mannose pyrophosphorylase B.

[0245]   Transferase-related sequences possess or interact with UDP-glucuronosyl and UDP-glucosyl transferase domains, which are protein domains found in a superfamily of enzymes that catalyze the addition of the glycosyl group from a UTP-sugar to a small hydrophobic molecule (http://pfam.wustl.edu/cgi-bin/getdesc?name=UDPGT). Transferase-related sequences also possess or interact with nucleotide transferase (NTP_transferase) domains, which are protein domains that transfer nucleotides onto phosphorylated sugars (http://pfam.wustl.edu/cgi-bin/getdesc?name=NTP_transferase).

Transposase-Related Sequences

[0246]   Transposases are site-specific recombination enzymes that catalyze the transposition of a segment of DNA from one part of the genome to another. The movable segments are called transposable elements; each transposable element is occasionally moved by a transposase, which functions as an integrase, by inserting DNA sequences into other DNA sequences. Transposases are often encoded by the DNA of the transposable element itself. Transposases bind specifically to terminal inverted repeats of 10-500 bp that are characteristically part of transposable elements (Smit and Riggs, 1996). They catalyze both cutting and pasting of a transposable element from one segment of the genome to another. Sequences related to transposases can have other functions, e.g., as transcription factors, or in the assembly of centromere proteins (Smit and Riggs, 1996). Examples of transposase-related sequences include mariner, pogo, hobo, tigger, MER37, Galileo, Occan, Impala, Tn MERI1, MsqTc3, and the sleeping beauty transposon system (Robertson and Zumpano, 1997; Robertson, 1996; Smit and Riggs, 1996).

[0247]   Transposase-related sequences can possess or interact with a transposase 1 (Transposase_1) domain, which is characterized by sequences that can excise and/or insert mobile genetic elements such as transposons or insertion sequences; for example, mariner possesses a transposase 1 domain (http://pfam.wustl.edu/cgi-bin/getdesc? name= Transposase_1). Transposase-related sequences can also possess or interact with L1 transposable element (Transposase_22) domains, which have been described above. Transposase-related sequences can also possess or interact with a DDE endonuclease (DDE) domain, which is responsible for coordinating metal ions needed for endonuclease catalytic activity (http://pfam.wustl.

edu/cgi-bin/getdesc? name=DDE). Transposase-related sequences can additionally possess or interact with a zinc finger, C2H2 type (zf-C2H2) domain, which bind nucleic acids using a mechanism that involves coordinating a zinc atom with a pair of cysteine residues and a pair of histidine residues (http://pfam.wustl.edu/cgi-bin/getdesc?name=zf-C2H2). Transposase-related sequences can also possess or interact with a reverse transcriptase (rvt) domain, and/or a low-density lipoprotein receptor (ldl_rece) domain, both of which are described above.

Ubiquitin-Related Sequences

[0248]   Ubiquitin is a protein found in all eucaryotic cells examined to date. When it is linked to the lysine side chain of a protein by the formation of an amide bond with its C-terminal glycine, ubiquitin renders the ubiquitin-bound protein subject to rapid proteolysis in the proteasome. In addition to its role in the selective degradation of cellular proteins, ubiquitin also plays a role in maintaining chromosome structure, regulating gene expression, responding to stresses on the organism, the regulation of gene expression, and ribosome biogenesis. Examples of ubiquitin-related sequences include elongins, ubiquitin-specific proteases, ubiquitin-calmodulin ligase, ubiquitin carrier protein kinase, ubiquitin N-alpha-protein hydrolase, and the small ubiquitin-related modifier (Sumo-1) (Kamitani et al., 1997).

[0249]   Ubiquitin-related sequences can possess or interact with a ubiquitin domain, which is a conserved sequence of approximately 76 amino acid residues that comprise the protein ubiquitin (http://pfam.wustl.edu/cgi-bin/getdesc?name=ubiquitin). Ubiquitin-related sequences can also possess or interact a ubiquitin carboxyl-terminal hydrolase (UCH) domain, which is a protein domain that comprises a thiol protease that recognizes and hydrolyses the peptide bond at the C-terminal glycine of ubiquitin (http://pfam.wustl.edu/cgi-bin/get desc?name=UCH).

Virus-Related Sequences

[0250]   The human chromosome has integrated endogenous genes that are related to viral genes. Some endogenous viral genes, e.g., the retroviral HERV-W family, are widely and heterogeneously dispersed among human chromosomes (Voisset et al., 2000; Everett et al., 1997; Werner et al., 1990). Endogenous proviruses are usually transcriptionally silent, but are expressed under certain conditions (Coffin et al., 1997). Endogenous viral expression can be specific to host factors, such as cell type or stage of differentiation, as well as other factors including

the position on the chromosome, the influence of *cis*-acting sequences, or the presence of host-mediated DNA methylation (Coffin).

[0251]　　Endogenous viral expression can have a number of consequences, both beneficial and detrimental. Among the beneficial consequences is the ability of endogenous retroviruses to confer resistance to infection by exogenous viruses. For example, mice with endogenous mouse mammary tumor virus (MMTV) can be immune to exogenous infection (Golovkina, et al., 1992). Among the detrimental effects is a causative role in disease. Evidence indicates an association between endogenous viruses with cancers and autoimmune diseases (Coffin et al., 1997). For example, spontaneous tumors of specific origin, murine mammary adenocarcinomas, and murine T-cell lymphomas have been associated with the presence of specific endogenous retroviruses. Furthermore, a transformed phenotype is associated with the increased transcription of certain classes of endogenous viral elements (Coffin et al., 1997). With respect to autoimmune disease, an endogenous virus that influences the immunoregulatory process has been associated with spontaneous autoimmune thyroiditis in a chicken model of human Hashimoto disease (Wick et al., 1987). Examples of viral-related proteins include hepatitis B virus x-interacting protein, herpesvirus associated ubiquitin-specific protease, and Coxsackievirus and adenovirus receptor precursor.

[0252]　　Viral-related sequences can possess or interact with rvt, rve, and gag_p30 sequences, all of which are described above.

Zinc Finger-Related Sequences

[0253]　A zinc finger domain is a small, self-folding, structural motif of 25 to 30 amino-acid residues present in many nucleic acid-binding proteins. It is comprised of a polypeptide loop held in a hairpin bend and bound to a zinc atom, and includes two conserved cysteine and two conserved histidine residues. Many classes of zinc fingers have been characterized according to the number and positions of the conserved histidine and cysteine residues. The amino acid configuration that holds the zinc atom in a tetrahedral array has a finger-like projection that interacts with nucleotides in the major groove of the bound nucleic acid. Zinc finger motifs have conserved regions near the zinc molecule, and variable regions at the nucleic acid binding site that provide specificity for the nucleic acid sequences they bind. Zinc finger proteins have a variety of functions, including as transcription regulators and intracellular receptors. Zinc finger domains are also involved in protein-protein

interactions, e.g., those involving protein kinase C. Recently, zinc finger nucleases have been used to target genes for gene replacement by homologous recombination (Bibikova et al., 2003). Examples of zinc finger proteins include XC3H-3b, the transcription factor Slug, and transcription factor IIIA.

[0254]   Zinc finger-related sequences can possess or interact with a zinc finger C2H2 type (zf-C2H2) domain, which binds a zinc atom with two cysteine and two histidine residues, and is utilized, e.g., in RNA transcription (http://pfam.wustl. edu/cgi-bin/getdesc?name=zf-C2H2). Zinc finger-related sequences can also possess or interact with a C3HC4 type, RING finger (zf-C3HC4) domain, which is a specialized type of zinc finger domain comprised of 40 to 60 amino acids that binds two zinc atoms; variants of RING-finger domains include the C3HC4-type and the C3H2C3-type (http://pfam.wustl.edu/cgi-bin/getdesc?name=zf-C3HC4). Proteins with RING-finger domains have developmental and functional roles; they are involved in intracellular receptor binding, and in mediating protein-protein interactions (Gray et al., 2000). RING-finger domains can exhibit ubiquitin-protein ligase activity, and can bind to E2 ubiquitin-conjugating enzymes.

[0255]   Zinc finger-related sequences can also possess or interact with a zinc knuckle (zf-CCHC) domain, which is an 18-amino acid zinc finger domain found in RNA-binding and single strand DNA-binding proteins; they are often involved in eukaryotic gene regulation (http://pfam.wustl.edu/cgi-bin/getdesc?name=zf-CCHC). Zinc knuckles are also found in retroviral gag and nucleocapsid proteins, where they function in genome packaging, and early in the infection process. Zinc finger-related sequences can also possess or interact with a BTB/POZ (BTB) domain, which mediates both homomeric and heteromeric protein dimerization (http://pfam.wustl. edu/cgi-bin/getdesc?name=BTB). Zinc finger-related sequences can also possess or interact with NF-X1 type zinc finger (zf-NF-X1) domains, which are found in the transcriptional repressor NK-X1, where they repress transcription of HLA-DRA, and in the shuttle craft protein, which plays a role in late stage embryonic neurogenesis (http://pfam.wustl.edu/cgi-bin/getdesc?name=zf-NF-X1). Zinc finger-related sequences can also possess or interact with a KRAB box (KRAB) domain, also known as a Kruppel-associated box, which is comprised of approximately 75 amino acids, enriched in charged amino acids, and involved in protein-protein interactions (http://pfam.wustl.edu/cgi-bin/getdesc? name=KRAB). KRAB domains can function as transcription factors, e.g., as a transcriptional repressor, and can assume roles in

cell differentiation and development (Aubry et al., 1992; Lovering and Trowsdale, 1991). Zinc finger-related sequences can possess or interact with a transposase_22 domain, which is described above.

INDUSTRIAL APPLICABILITY

[0256]   The invention provides sequences related to secreted sequences, single-transmembrane sequences, multiple-transmembrane sequences, kinase-related sequences, ligase-related sequences, nuclear hormone receptor-related sequences, phosphatase-related sequences, protease-related sequences, phosphodiesterase-related sequences, kinesin-related sequences, immunoglobulin-related sequences, T-cell receptor-related sequences, glycosylphosphatidylinositol anchor-related sequences, and sequences related to other nucleic acid and amino acid sequences of the invention, including activators, adaptors, adhesion molecules, ATPases, ATP, breakpoints, channels, checkpoints, complexes, dehydrogenases, disintegrins, endopeptidases, germ-cells, GTPases, helicases, hydrolases, integrases, integrins, isomerases, membranes, mucins, oxygenases, peroxidases, phopholipases, prosaposins, proteosomes, reductases, reverse trancriptases, RNases, RNases H, SH3, synthetases, TATA boxes, Tat proteins, transferases, transposases, ubiquitins, and viruses. The invention provides for novel polynucleotides, related novel polypeptides and active fragments thereof, as well as novel nucleic acid compositions encoding these polypeptides, compositions comprising the related polypeptides, and methods for their use.

[0257]   The present invention also provides for vectors, host cells, and methods for producing the polynucleotides and polypeptides of the invention in these vectors and host cells. The present invention further provides for antisense molecules that are capable of regulating the expression of the polynucleotides or polypeptides herein. In addition, modulators, including antibodies that bind specifically to the polypeptides or modulate the activity of the polypeptides, are also provided.

[0258]   The present polynucleotides, polypeptides, and modulators find use in therapeutic agent screening/discovery applications, such as screening for receptors or competitive ligands, for use, for example, as small molecule therapeutic drugs. Also provided are methods of modulating a biological activity of a polypeptide and methods of treating associated disease conditions, particularly by administering

modulators of the present polypeptides, such as small molecule modulators, antisense molecules, and specific antibodies.

[0259]   The present polypeptides, polynucleotides, and modulators find use in a number of diagnostic, prophylactic, and therapeutic applications. The polynucleotides and polypeptides of the invention can be detected by methods provided herein; these methods are useful in diagnosis, and can be accomplished by the use of diagnostic kits. The polynucleotides and polypeptides of the invention are useful for treating a variety of disorders, including cancer, proliferative disorders, inflammatory disorders, immune disorders, bacterial or viral disorders, and other metabolic disorders. For example, subjects who suffer from a deficiency, or a lack of a particular protein, or are otherwise in need of such protein to repair or enhance a desirable function, benefit from the administration of a protein or an active fragment thereof by any conventional routes of administration. These include therapeutic vaccines in the form of nucleic acid or polypeptide vaccines, such as cancer vaccines, where the vaccines can be administered alone, such as naked DNA, or can be facilitated, such as via viral vectors, microsomes, or liposomes. Therapeutics antibodies include those that are administered alone or in combination with cytotoxic agents, such as radioactive or chemotherapeutic agents.

[0260]   In particular, the polypeptides, polynucleotides, and modulators of the present invention can be used to treat cancers, including, but not limited to, cancers of the prostate, breast, bone, soft tissue, liver, kidney, ovary, cervix, skin, pancreas, and brain, as well as leukemias, lymphomas, lung cancers such as adenocarcinomas and squamous cell carcinoma, and cancers of gastrointestinal organs such as stomach, colon, and rectum. Further, the polypeptides, polynucleotides, and modulators of the present invention can be used to treat inflammatory, immune, bacterial, viral, and metabolic diseases, disorders, syndromes, or conditions, including, but not limited to, intestinal inflammation and immunity, autoimmune thyroiditis, and retroviral infections, as well as tissue and/or organ hypertrophy.

DISCLOSURE OF THE INVENTION

[0261]   The present invention features an isolated polynucleotide that encodes a polypeptide. In some embodiments, the polypeptide has at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 97%, at least about 98%, or at least about 99% amino acid

sequence identity with an amino acid sequence derived from a polynucleotide
sequence chosen from at least one nucleotide sequence according to SEQ ID NOS.: 1
- 1231 and 2463 - 3697. In some embodiments, the polypeptide has an amino acid
sequence chosen from at least one amino acid sequence according to SEQ ID. NOS.
1232 - 2462. In many embodiments, the polypeptide has at least one activity
associated with the naturally occurring encoded polypeptide.

[0262]   In some embodiments, the polypeptide includes a signal peptide. In
alternative embodiments, the polypeptide comprises a mature form of a protein, from
which the signal peptide has been cleaved. In other embodiments, the polypeptide is a
signal peptide. In a further aspect, the invention provides fragments of a polypeptide
chosen from at least one amino acid sequence according to SEQ ID NOS.: 1232 -
2462, where each fragment is an extracellular fragment of the polypeptide, or an
extracellular fragment of the polypeptide minus the signal peptide. The invention
provides an N-terminal fragment containing a Pfam domain and a C-terminal
fragment containing a Pfam domain and either or both may be biologically active.

[0263]   In yet other embodiments, the polypeptides function as secreted
proteins. In yet further embodiments, the polypeptides function as single-
transmembrane proteins. In yet further embodiments, the polypeptides function as
multiple-transmembrane proteins. In yet further embodiments, the polypeptides
function as kinases. In yet further embodiments, the polypeptides function as protein
kinases. In yet further embodiments, the polypeptides function as ligases. In yet
further embodiments, the polypeptides function as nuclear hormone receptors. In yet
further embodiments, the polypeptides function as phosphatases. In yet further
embodiments, the polypeptides function as proteases. In yet further embodiments, the
polypeptides function as phosphodiesterases. In yet further embodiments, the
polypeptides function as kinesins. In yet further embodiments, the polypeptides
function as immunoglobulins. In yet further embodiments, the polypeptides function
as T-cell receptors. In yet further embodiments, the polypeptides function as
glycosylphosphatidylinositol anchors.

[0264]   In yet further embodiments, the polypeptides function as cytokines.
In still further embodiments, the polypeptides function as immune cells. In further
embodiments, the polypeptides function as antigens. In yet further embodiments, the
polypeptides function as receptors. In other embodiments, the polypeptides function
as binding proteins. In other embodiments, the polypeptides function as factors. In

further embodiments, the polypeptides function as growth factors. In further embodiments, the polypeptides function as heat-shock proteins. In some embodiments, the polypeptides function as membrane transport proteins. In yet further embodiments, the polypeptides function as ribosomal proteins. In some embodiments, the polypeptides function as zinc fingers. In some embodiments, the polypeptides function as embryonic stem cell-related peptides. In still further embodiments, the polypeptides function in pathological states. In other embodiments, the polypeptides function as one or more of these.

[0265] In yet further embodiments, the polypeptides function as activators. In yet further embodiments, the polypeptides function as adaptors. In yet further embodiments, the polypeptides function as adhesion molecules. In yet further embodiments, the polypeptides function as ATPases. In yet further embodiments, the polypeptides function as ATP-related polypeptides. In further embodiments, the polypeptides function as channel-related polypeptides. In yet further embodiments, the polypeptides function as checkpoint-related polypeptides. In yet further embodiments, the polypeptides function as complexes. In yet further embodiments, the polypeptides function as dehydrogenases. In yet further embodiments, the polypeptides function as disintegrins. In yet further embodiments, the polypeptides function as endopeptidases. In yet further embodiments, the polypeptides function as germ-cells. In yet further embodiments, the polypeptides function as GTPases. In yet further embodiments, the polypeptides function as helicases. In yet further embodiments, the polypeptides function as hydrolases. In yet further embodiments, the polypeptides function as integrases. In yet further embodiments, the polypeptides function as integrins. In yet further embodiments, the polypeptides function as isomerases. In yet further embodiments, the polypeptides function as membranes. In yet further embodiments, the polypeptides function as mucins. In yet further embodiments, the polypeptides function as oxygenases. In yet further embodiments, the polypeptides function as peroxidases. In some embodiments, the polypeptides function as phospholipases. In yet further embodiments, the polypeptides function as prosaposins. In yet further embodiments, the polypeptides function as proteasomes. In yet further embodiments, the polypeptides function as reductases. In other embodiments, the polypeptides function as reverse transcriptase-related polypeptides. In yet further embodiments, the polypeptides function as RNases. In further embodiments, the polypeptides function as RNase H-related polypeptides. In yet

further embodiments, the polypeptides function as SH3-related polypeptides. In yet further embodiments, the polypeptides function as synthetases. In yet further embodiments, the polypeptides function as TATA box-related polypeptides. In yet further embodiments, the polypeptides function as TAT-related polypeptides. In yet further embodiments, the polypeptides function as transferases. In yet further embodiments, the polypeptides function as transposases. In yet further embodiments, the polypeptides function as ubiquitin-related polypeptides. In yet further embodiments, the polypeptides function as virus-related polypeptides. In other embodiments, the polypeptides function as one or more of these.

[0266] The present invention features an isolated polynucleotide that hybridizes under stringent hybridization conditions to a coding region of at least one nucleotide sequence shown in SEQ ID NOS.: 1 - 1231, 2463 - 3697, or a complement thereof.

[0267] The present invention features an isolated polynucleotide that shares at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 97%, at least about 98%, at least about 99% nucleotide sequence identity with a nucleotide sequence of the coding region of at least one sequence shown in SEQ ID NOS.: 1 - 1231, 2463 - 3697, or a complement thereof. In some embodiments, a subject polynucleotide has the nucleotide sequence shown in at least one of SEQ ID NOS.: 1 - 1231, 2463 - 3697, or a coding region thereof.

[0268] The present invention also features a vector, e.g., a recombinant vector, that includes a subject polynucleotide, and a promoter the drives its expression. This vector can transform a host cell, and the present invention further features such host cells, e.g., isolated *in vitro* host cells, and *in vivo* host cells, that comprise a polynucleotide of the invention, or a recombinant vector of the invention.

[0269] The present invention further features a library of polynucleotides, wherein at least one of the polynucleotides comprises the sequence information of a polynucleotide of the invention. In specific embodiments, the library is provided on a nucleic acid array. In some embodiments, the library is provided in computer-readable format.

[0270] The present invention features a pair of isolated nucleic acid molecules, each from about 10 to about 200 nucleotides in length. The first nucleic acid molecule of the pair comprises a sequence of at least 10 contiguous nucleotides

having 100% sequence identity to at least one nucleic acid sequence shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697. The second nucleic acid molecule of the pair comprises a sequence of at least 10 contiguous nucleotides having 100% sequence identity to the reverse complement of at least one nucleic acid sequence shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697. The sequence of said second nucleic acid molecule is located 3' of the nucleic acid sequence of the first nucleic acid molecule shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697. The pair of isolated nucleic acid molecules are useful in a polymerase chain reaction or in any other method known in the art to amplify a nucleic acid that has sequence identity to the sequences shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697, particularly when cDNA is used as a template.

[0271] The invention features a method of determining the presence of a polynucleotide substantially identical to a polynucleotide sequence shown in the Sequence Listing, or a complement of such a nucleotide by providing its complement, allowing the polynucleotides to interact, and determining whether such interaction has occurred.

[0272] The invention further features methods of regulating the expression of the subject polynucleotides and encoded polypeptides. The invention provides a method of inhibiting transcription or translation of a first polynucleotide encoding a first polypeptide of the invention by providing a second polynucleotide that hybridizes to the first polynucleotide, and allowing the first polynucleotide to contact and bind to the second polynucleotide. The second polynucleotide can be chosen from an antisense molecule, a ribozyme, and an interfering RNA (RNAi) molecule.

[0273] The present invention further features an isolated polypeptide, e.g., an isolated polypeptide encoded by a polynucleotide, and biologically active fragments of such polypeptide. In some embodiments, the polypeptide is a fusion protein. In some embodiments, the polypeptide has one or more amino acid substitutions, and/or insertions and/or deletions, compared with at least one sequence shown in SEQ ID NOS.: 1232 - 2462. In some embodiments, the polypeptide has an amino acid sequence derived from at least one nucleotide sequence shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697. In some embodiments, the polypeptide has an amino acid sequence substantially identical to at least one sequence shown in SEQ ID NOS.: 1232 - 2462.

[0274] The invention also provides a method of making a polypeptide of the invention by providing a nucleic acid molecule that comprises a polynucleotide sequence encoding a polypeptide of the invention, introducing the nucleic acid molecule into an expression system, and allowing the polypeptide to be produced.

[0275] In some embodiments, the method involves *in vitro* cell-free transcription and/or translation. For example, the expression system can comprise a cell-free expression system, such as an *E. coli* system, a wheat germ extract system, a rabbit reticulocyte system, or a frog oocyte system.

[0276] In certain other embodiments, the expression system can comprise a prokaryotic or eukaryotic cell, for example, a bacterial cell expression system, a fungal cell expression system, such as yeast or *Aspergillus*, a plant cell expression system, e.g., a cereal plant, a tobacco plant, a tomato plant, or other edible plant, an insect cell expression system, such as SF9 of High Five cells, an amphibian cell expression system, a reptile cell expression system, a crustacean cell expression system, an avian cell expression system, a fish cell expression system, or a mammalian cell expression system, such as one using Chinese Hamster Ovary (CHO) cells. In some embodiments, the method involves culturing a subject host cell under conditions such that the subject polypeptide is produced by the host cells; and recovering the subject polypeptide from the culture, e.g., from within the host cells, or from the culture medium. In further embodiments, the polypeptide can be produced *in vivo* in a multicellular animal or plant, comprising a polynucleotide encoding the subject polypeptide.

[0277] The present invention further features a non-human animal injected with at least one polynucleotide comprising at least one nucleotide sequence chosen from SEQ ID NOS.: 1 - 1231 and 2463 - 3697, and/or at least one polypeptide comprising at least one amino acid sequence chosen form SEQ ID NOS.: 1232 - 2462.

[0278] The present invention further features an antibody that specifically recognizes, binds to, interferes with, or modulates the biological activity of a subject polypeptide or a fragment thereof. The polypeptide can be a single-transmembrane protein, multiple-transmembrane protein, kinase, protein kinase, ligase, nuclear hormone receptor, phosphatase, protease, phosphodiesterase, kinesin, immunoglobulin, T-cell receptor, glycosylphosphatidylinositol anchor, or other nucleic acid and amino acid sequences, including, activators, adaptors, adhesion molecules, ATPases, ATP, breakpoints, channels, checkpoints, complexes,

dehydrogenases, disintegrins, endopeptidases, germ-cells, GTPases, helicases, hydrolases, integrases, integrins, isomerases, membranes, mucins, oxygenases, peroxidases, phospholipases, prosaposins, proteasomes, reductases, reverse transcriptases, RNases, RNases H, SH3, synthetases, TATA boxes, Tat, transferases, transposases, ubiquitins, and viruses. The fragment can be an extracellular fragment of a subject polypeptide, or an extracellular fragment of a subject polypeptide minus the signal peptide.

[0279] The present invention further features an antibody that specifically inhibits binding of a polypeptide to its ligand or substrate. It also features an antibody that specifically inhibits binding of a polypeptide as a substrate to another molecule.

[0280] Another aspect of the present invention features a library of antibodies or fragments thereof, wherein at least one antibody or fragment thereof specifically binds to at least a portion of a polypeptide comprising an amino acid sequence according to SEQ ID NOS.: 1232 - 2462 , and/or wherein at least one antibody or fragment thereof interferes with at least one activity of such polypeptide or fragment thereof. In certain embodiments, the antibody library comprises at least one antibody or fragment thereof that specifically inhibits binding of a subject polypeptide to its ligand or substrate, or that specifically inhibits binding of a subject polypeptide as a substrate to another molecule. The present invention also features corresponding polynucleotide libraries comprising at least one polynucleotide sequence that encodes an antibody or antibody fragment of the invention. In specific embodiments, the library is provided on a nucleic acid array or in computer-readable format.

[0281] An antibody of the present invention may comprise a monoclonal antibody, polyclonal antibody, single chain antibody, intrabody, and active fragments of any of these. The active fragments include variable regions from either heavy chains or light chains. The antibody can comprise the backbone of a molecule with an immunoglobulin domain, e.g., a fibronectin backbone, a T-cell receptor backbone, or a CTLA4 backbone.

[0282] The present invention further features a targeting antibody, a neutralizing antibody, a stabilizing antibody, an enhancing antibody, an antibody agonist, an antibody antagonist, an antibody that promotes cellular endocytosis of a target antigen, a cytotoxic antibody, and an antibody that mediates antibody dependent cellular cytotoxicity (ADCC). The antibody that mediates ADCC can have

a cytotoxic component, e.g., a radioisotope, a radioactive molecule, a microbial toxin, a plant toxin, a chemotherapeutic agent, or a chemical substance, such as doxorubicin or cisplatin. The invention also features an inhibitory antibody, functioning to specifically inhibit the binding of a cognate polypeptide to its ligand or its substrate, or to specifically inhibit the binding of a cognate peptide as the substrate of another molecule.

[0283] The antibodies of the present invention also encompass a human antibody, a non-human primate antibody, a monkey antibody, a non-primate animal antibody, e.g., a rodent antibody, rat antibody, a mouse antibody, a hamster antibody, a guinea pig antibody, a chicken antibody, a cattle antibody, a sheep antibody, a goat antibody, a horse antibody, porcine antibody, a cow antibody, a rabbit antibody, a cat antibody, or a dog antibody. It also features a humanized antibody, a primatized antibody, and a chimeric antibody.

[0284] The antibodies of the invention can be produced *in vitro* or *in vivo*. For example, the present invention features an antibody produced in a cell-free expression system, a prokaryote expression system or a eukaryote expression system, as described herein.

[0285] The invention further provides a host cell that can produce an antibody of the invention or a fragment thereof. The antibody may also be secreted by the cell. The host cell can be a hybridoma, or a prokaryotic or eukaryotic cell. The invention also provides a bacteriophage or other virus particle comprising an antibody of the invention, or a fragment thereof. The bacteriophage or other virus particle may display the antibody or fragment thereof on its surface, and the bacteriophage itself may exist within a bacterial cell. The antibody may also comprise a fusion protein with a viral or bacteriophage protein.

[0286] The invention further provides transgenic multicellular organisms, e.g., plants or non-human animals, as well as tissues or organs, comprising a polynucleotide sequence encoding a subject antibody or fragment thereof. The organism, tissues, or organs will generally comprise cells producing an antibody of the invention, or a fragment thereof.

[0287] In another aspect, the present invention features a method of making an antibody by immunizing a host animal. In this method, a polypeptide or a fragment thereof, a polynucleotide encoding a polypeptide, or a polynucleotide encoding a fragment thereof, is introduced into an animal in a sufficient amount to

elicit the generation of antibodies specific to the polypeptide or fragment thereof, and the resulting antibodies are recovered from the animal. The polypeptide can be encoded by a nucleic acid molecule comprising a nucleotide sequence chosen from at least one polynucleotide sequence according to SEQ ID NOS.: 1 - 1231 and 2463 - 3697. For example, the polypeptide may comprise at least one amino acid sequence chosen from SEQ ID NOS.: 1232 - 2462 .

[0288] The invention thus also provides a non-human animal comprising an antibody of the invention. The animal can be a non-human primate, (e.g., a monkey) a rodent (e.g., a rat, a mouse, a hamster, a guinea pig), a chicken, cattle (e.g., a sheep, a goat, a horse, a pig, a cow), a rabbit, a cat, or a dog.

[0289] The present invention also features a method of making an antibody by isolating a spleen from an animal injected with a polypeptide or a fragment thereof, a polynucleotide encoding a polypeptide, or a polynucleotide encoding a fragment thereof, and recovering antibodies from the spleen cells. Hybridomas can be made from the spleen cells, and hybridomas secreting specific antibodies can be selected.

[0290] The present invention further features a method of making a polynucleotide library from spleen cells, and selecting a cDNA clone that produces specific antibodies, or fragments thereof.. The cDNA clone or a fragment thereof can be expressed in an expression system that allows production of the antibody or a fragment thereof, as provided herein.

[0291] The invention also provides a method for determining the presence or measuring the level of a polypeptide that specifically binds to an antibody of the invention. This method involves allowing the antibody to interact with a sample, and determining whether interaction between the antibody and any polypeptide in the sample has occurred. Antibodies that specifically bind to at least one subject polypeptide are useful in diagnostic assays, e.g., to detect the presence of a subject polypeptide. Similarly, the invention features a method of determining the presence of an antibody to a polypeptide of the invention, by providing the polypeptide, allowing the antibody and the polypeptide to interact, and determining whether interaction has occurred.

[0292] The present invention further features a method of identifying an agent that modulates the level of a subject polypeptide (or an mRNA encoding a subject polypeptide) in a cell. The method generally involves contacting a cell (e.g., a

eukaryotic cell) that produces the subject polypeptide with a test agent; and determining the effect, if any, of the test agent on the level of the polypeptide in the cell.

[0293]   The present invention further features a method of identifying an agent that modulates biological activity of a subject polypeptide. The methods generally involve contacting a subject polypeptide with a test agent; and determining the effect, if any, of the test agent on the activity of the polypeptide. In certain embodiments, the polypeptide is expressed on a cell surface. In certain embodiments, the agent or modulator is an antibody, for example, where an antibody binds to the polypeptide or affects its biological activity.

[0294]   The present invention further features biologically active agents (or modulators) identified using a method of the invention.

[0295]   The present invention also features a method of modulating biological activity using an agent selectable by the above methods. Briefly, the method of modulating biological activity comprises contacting the agent with a first human or a non-human host cell, thereby modulating the activity of the first host cell or a second host cell. In one example, contacting the agent with the first human or non-human host cell results in the recruitment of a second host cell. The agent may be an antibody or antibody fragment of the invention.

[0296]   The modulation can comprise directly enhancing cell activity, indirectly enhancing cell activity, directly inhibiting cell activity, or indirectly inhibiting cell activity. The cell activity that is modulated can include transcription, translation, cell cycle control, signal transduction, intracellular trafficking, cell adhesion, cell mobility, proteolysis, ion transport, water transport, DNA repair, hydrolysis, lipase activity, polymerization using an RNA temple or a DNA template, and nuclease activity. The modulation can result in cell death or apoptosis, or inhibition of cell death or apoptosis, as well as cell growth, cell proliferation, or cell survival, or inhibition of cell growth, cell proliferation, or cell survival; as well as mucosal preservation, inhibition of eicosanoid synthesis, or resistance to infection by viruses.

[0297]   Either the first or the second host cell can be a human or a non-human host cell. Either the first or the second host cell can be an immune cell, e.g., a T cell, B cell, NK cell, dendritic cell, macrophage, muscle cell, stem cell, skin cell, fat cell, blood cell, brain cell, bone marrow cell, endothelial cell, retinal cell, bone cell,

kidney cell, pancreatic cell, liver cell, spleen cell, prostate cell, cervical cell, ovarian cell, breast cell, lung cell, liver cell, soft tissue cell, colorectal cell, other cell of the gastrointestinal tract, or a cancer cell.

[0298] The invention also provides a method of diagnosing cancer, proliferative, inflammatory, immune, viral, bacterial, or metabolic disorder in a patient, by allowing an antibody specific for a polypeptide of the invention to contact a patient sample, and detecting specific binding between the antibody and any antigen in the sample to determine whether the subject has cancer, proliferative, inflammatory, immune, viral, bacterial, or metabolic disorder.

[0299] The invention further provides a method of diagnosing cancer, proliferative, inflammatory, immune, viral, bacterial, or metabolic disorder in a patient, by allowing a polypeptide of the invention to contact a patient sample, and detecting specific binding between the polypeptide and any interacting molecule in the sample to determine whether the subject has cancer, proliferative, inflammatory, immune, viral, bacterial, or metabolic disorder.

[0300] The invention also features a method of providing a polynucleotide, a polypeptide, or an agent of the invention, such as an antibody, to a subject by oral, buccal, nasal, rectal, intraperitoneal, intradermal, transdermal, intratracheal, intrathecal, or parenteral administration, or otherwise by implantation or inhalation. For example, the polynucleotide, polypeptide or agent can be administered intranasally, intravenously, intra-arterially, intracardiacally, subcutaneously, intraperitoneally, transdermally, intraventricularly, or intracranially. The invention also provides a method for formulating a polynucleotide, polypeptide, or modulator composition, such as an antibody composition, for delivery by any of the routes of administration provided above, for example, for treatment of disorders. For example, the parenteral delivery can be via inhalation or implantation. The parenteral delivery can also be oral, intranasal, intraventricular, or intracranial.

[0301] The present invention also features a pharmaceutical composition comprising a polynucleotide, polypeptide, or modulator of the invention and a carrier. The carrier can be a pharmaceutically acceptable carrier. The modulator can be obtainable by any methods of the invention, for example, the modulator can be an antibody or a fragment thereof. Further, oral formulations, preparations for injection, aerosol formulations, and suppositories can be prepared, each comprising the polynucleotide, polypeptide, or modulator composition. Further, nucleic acid

compositions comprising polynucleotide sequences encoding the subject antibodies, or fragments thereof, can be prepared for administration to a subject.

[0302]   The invention also features a non-human animal injected with the polynucleotide, polypeptide, or modulator composition, for example the antibody composition. Again, the animal can be a non-human primate, (e.g., a monkey) a rodent (e.g., a rat, a mouse, a hamster, a guinea pig), a chicken, cattle (e.g., a sheep, a goat, a horse, a pig, a cow), a rabbit, a cat, or a dog.

[0303]   In another aspect, the invention provides a method of treating a disorder in a subject needing or desiring such treatment, comprising administering a polynucleotide, polypeptide, or modulator of the invention to the subject. The subject can be a human or a non-human animal. The disorder can be cancer, proliferative, inflammatory, immune, metabolic, ulcerative, bacterial, or viral disorders.

[0304]   For example, the method of treatment may comprise administering an antibody composition with a first antibody that specifically binds to a first epitope of a first polypeptide or a fragment thereof, or that interferes with at least one activity of the first polypeptide or a fragment thereof, wherein the first polypeptide is encoded by a nucleic acid molecule comprising a nucleotide sequence chosen from SEQ ID NOS.: 1 - 1231 and 2463 - 3697, or any nucleic acid of the present invention. For example, the first polypeptide may comprise an amino acid sequence chosen from SEQ ID NOS.: 1232 - 2462. In certain embodiments, this method further comprises using a second antibody that binds specifically to or interferes with the activity of a second epitope of the first polypeptide or to a first epitope of a second polypeptide. The second polypeptide can be encoded by a nucleic acid molecule comprising a nucleotide sequence chosen from SEQ ID NOS.: 1 - 1231 and 2463 - 3697, or any nucleic acid of the present invention. For example, the second polypeptide may comprise an amino acid sequence chosen from SEQ ID NOS.: 1232 - 2462. In certain embodiments, the antibody binds, or interferes with the activity of, at least one polypeptide fragment, wherein the fragment is an extracellular fragment of the polypeptide, or an extracellular fragment of the polypeptide minus the signal peptide, for the treatment, for example, of proliferative disorders, such as cancer.

[0305]   In other embodiments, the modulator may bind to a cell surface molecule that is over-expressed in the disorder. Further the modulator may be linked to an antibody of the invention. The antibody can be capable of initiating antibody dependent cell cytotoxicity, e.g., where the antibody is in turn coupled to cytotoxic

agents. This method is applicable when the disorder is cancer, another proliferative disorder, inflammatory, immune, bacterial, viral, or metabolic disorder, and the cell surface molecule is over-expressed in a cancer cell, diseased cell or virus-infected cell. The cell surface molecule can be a single-transmembrane-related protein, a multiple-transmembrane-related protein, a kinase-related protein, a protein kinase-related protein, a ligase-related protein, a nuclear hormone receptor-related protein, a phosphatase-related protein, a protease-related protein, a phosphodiesterase-related protein, a kinesin-related protein, an immunoglobulin-related protein, a T-cell receptor-related protein, a glycosylphosphatidylinositol anchor-related protein, or other amino acid sequence, including, an activator-related protein, an adaptor-related protein, an adhesion molecule-related protein, an ATPase-related protein, an ATP-related protein, a breakpoint-related protein, a channel-related protein, a checkpoint-related protein, a complex-related protein, a dehydrogenase-related protein, a disintegrin-related protein, an endopeptidase-related protein, a germ-cell-related protein, a GTPase-related protein, a helicase-related protein, a hydrolase-related protein, an integrase-related protein, an integrin-related protein, isomerase-related protein, a membrane-related protein, a mucin-related protein, an oxygenase-related protein, a peroxidase-related protein, a phopholipase-related protein, a prosaposin-related protein, a proteasome-related protein, a reductase-related protein, a reverse transcriptase-related protein, an RNase-related protein, an RNase H-related protein, an SH3-related protein, a synthetase-related protein, a TATA box-related protein, a Tat-related protein, a transferase-related protein, a transposase-related protein, a ubiquitin-related protein, or virus-related protein that is over-expressed in cancer, proliferative, inflammatory, immune, bacterial, viral, or metabolic disorder.

[0306] The invention also provides a method for prophylactic or therapeutic treatment of a subject needing or desiring such treatment by providing a vaccine, that can be administered to the subject. The vaccine may comprise one or more of a polynucleotide, polypeptide, or modulator of the invention, for example an antibody vaccine composition, a polypeptide vaccine composition, or a polynucleotide vaccine composition, useful for treating cancer, proliferative, inflammatory, immune, metabolic, bacterial, or viral disorders.

[0307] For example, the vaccine can be a cancer vaccine, and the polypeptide can concomitantly be a cancer antigen. The vaccine may be an anti-inflammatory vaccine, and the polypeptide can concomitantly be an inflammation-

related antigen. The vaccine may be a viral vaccine, and the polypeptide can concomitantly be a viral antigen. In some embodiments, the vaccine comprises a polypeptide fragment, comprising at least one extracellular fragment of a polypeptide of the invention, and/or at least one extracellular fragment of a polypeptide of the invention minus the signal peptide, for the treatment, for example, of proliferative disorders, such as cancer. In certain embodiments, the vaccine comprises a polynucleotide encoding one or more such fragments, administered for the treatment, for example, of proliferative disorders, such as cancer. Further, the vaccine can be administered with or without an adjuvant.

[0308]　In another aspect, the invention provides a method for gene therapy by providing a polynucleotide comprising a nucleic acid molecule encoding a polypeptide, such as an antibody of the invention, and administering the polynucleotide to a subject needing or desiring such treatment.

[0309]　The invention further provides a kit comprising one or more of a polynucleotide, polypeptide, or modulator composition, such as an antibody composition, which may include instructions for its use. Such kits are useful in diagnostic applications, for example, to detect the presence and/or level of a polypeptide in a biological sample by specific antibody interaction.

MODES FOR CARRYING OUT THE INVENTION

Brief Description of the Tables

[0310]　Each sequence shown in Tables 1 and 2 is identified by a Five Prime Therapeutics, Inc. (FP) identification number (FP ID). Table 1 sets forth a profile of the expression patterns of some of the claimed sequences (Expression Profile), based on their homology with previously described sequences. This expression profile details the cell and tissue types that express the sequences, and the relative level of expression of the sequences. Table 1 specifies the predicted number of amino acid residues in each FP protein of the invention (Length, Predicted Protein). Table 1 also specifies the percent of the FP sequence that is covered by the public National Center for Information Biotechnology (NCBI) database (Prediction Covered by Public). Table 1 sets forth a profile of the subcellular localization of some of the claimed sequences, based on their homology with previously described sequences. (Subcellular Localization).

[0311]   Table 1 also describes the characteristics of the protein in the NCBI database displaying the greatest degree of similarity to each claimed sequence. This protein is described by its NCBI accession number (Top Hit Accession No.), and by the NCBI's annotation of that sequence (Top Hit Annotation). Finally, the predicted utilities of the claimed sequences, based on their homology with previously described sequences, are presented (Utility).

[0312]   Table 2 describes the characteristics of the human protein with the greatest degree of similarity to the claimed sequences present in the NCBI database. The predicted number of amino acids of this human protein is specified (Length, Human Top Hit). Table 2 also specifies any existing protein family (Pfam) classification for these human sequences. Table 2 specifies the result of the algorithm described above that predicts whether the claimed FP sequence is secreted (Tree Vote, Secreted). Table 2 sets forth the the position of the amino acid residues comprising the signal peptide sequences (SP Positions) of the claimed FP sequences. Table 2 also specifies the position(s), if any, of the amino acid residues comprising the transmembrane domains in each claimed FP sequence (TM domains), and the number of transmembrane domains of each claimed FP sequence (TM Total).

Definitions

[0313]   "Related sequences" include nucleotide and amino acid sequences that are involved in the function of their referent. For example, "receptor-related sequences" include all sequences that are involved in receptor function. This includes, but is not limited to, sequences that are involved in receptor synthesis, receptor regulation, receptor effector function, and receptor degradation. "Related sequences" also encompass complementary nucleic acid sequences, and biologically active fragments of nucleic acid and amino acid sequences.

[0314]   The terms "polynucleotide," "nucleotide," "nucleic acid," "polynucleic molecule," "nucleotide molecule," "nucleic acid molecule," "nucleic acid sequence," "polynucleotide sequence," and "nucleotide sequence" are used interchangeably herein to refer to polymeric forms of nucleotides of any length. The polynucleotides can contain deoxyribonucleotides, ribonucleotides, and/or their analogs or derivatives. For example, nucleic acids can be naturally occurring DNA or RNA, or can be synthetic analogs, as known in the art. The terms also encompass genomic DNA, genes, gene fragments, exons, introns, regulatory sequences or regulatory elements (such as promoters, enhancers, initiation and termination regions,

other control regions, expression regulatory factors, and expression controls), DNA comprising one or more single-nucleotide polymorphisms (SNPs), allelic variants, isolated DNA of any sequence, and cDNA. The terms also encompass mRNA, tRNA, rRNA, ribozymes, splice variants, antisense RNA, antisense conjugates, RNAi, and isolated RNA of any sequence. The terms also encompass recombinant polynucleotides, heterologous polynucleotides, branched polynucleotides, labeled polynucleotides, hybrid DNA/RNA, polynucleotide constructs, vectors comprising the subject nucleic acids, nucleic acid probes, primers, and primer pairs. The polynucleotides can comprise modified nucleic acid molecules, with alterations in the backbone, sugars, or heterocyclic bases, such as methylated nucleic acid molecules, peptide nucleic acids, and nucleic acid molecule analogs, which may be suitable as, for example, probes if they demonstrate superior stability and/or binding affinity under assay conditions. Analogs of purines and pyrimidines, including radiolabeled and fluorescent analogs, are known in the art. The polynucleotides can have any three-dimensional structure, and can perform any function, known or as yet unknown. The terms also encompass single-stranded, double-stranded and triple helical molecules that are either DNA, RNA, or hybrid DNA/RNA and that may encode a full-length gene or a biologically active fragment thereof. Biologically active fragments of polynucleotides can encode the polypeptides herein, as well as anti-sense and RNAi molecules. Thus, the full length polynucleotides herein may be treated with enzymes, such as Dicer, to generate a library of short RNAi fragments which are within the scope of the present invention.

[0315] The novel polynucleotides herein include those shown in the Tables, SEQ ID NOS.: 1 - 1231 and 2463 - 3697, as well as those that encode the polypeptides of SEQ ID NOS.: 1232 - 2462, and biologically active fragments thereof. The polynucleotides also include modified, labeled, and degenerate variants of the nucleic acid sequences, as well as nucleic acid sequences that are substantially similar or homologous to nucleic acids encoding the subject proteins.

[0316] A "biologically active" entity, or an entity having "biological activity," is one having structural, regulatory, or biochemical functions of a naturally occurring molecule or any function related to or associated with a metabolic or physiological process. Biologically active polynucleotide fragments are those exhibiting activity similar, but not necessarily identical, to an activity of a polynucleotide of the present invention. The biological activity can include an

improved desired activity, or a decreased undesirable activity. For example, an entity demonstrates biological activity when it participates in a molecular interaction with another molecule, or when it has therapeutic value in alleviating a disease condition, or when it has prophylactic value in inducing an immune response to the molecule, or when it has diagnostic value in determining the presence of the molecule, such as a biologically active fragment of a polynucleotide that can be detected as unique for the polynucleotide molecule, or that can be used as a primer in PCR.

[0317] The term "degenerate variant" of a nucleic acid sequence refers to all nucleic acid sequences that can be directly translated, according to the standard genetic code, to provide an amino acid sequence identical to that translated from a reference nucleic acid sequence.

[0318] The term "gene" or "genomic sequence" as used herein is an open reading frame encoding specific proteins and polypeptides, for example, an mRNA, cDNA, or genomic DNA, and also may or may not include intervening introns, or adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression up to about 20 kb beyond the coding region, and possibly further in either direction. A gene can be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome.

[0319] The term "transgene" as used herein is a nucleic acid sequence that is incorporated into a transgenic organism. A "transgene" can contain one or more transcriptional regulatory sequences, and other sequences, such as introns, that may be useful for expressing or secreting the nucleic acid or fusion protein it encodes.

[0320] The term "cDNA" as used herein is intended to include all nucleic acids that share the sequence elements of mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Generally, mRNA species have contiguous exons, the intervening introns having been removed by nuclear RNA splicing to create a continuous open reading frame encoding a protein.

[0321] The term "splice variant" refers to all types of RNAs transcribed from a given gene that when processed collectively encode plural protein isoforms. The term "alternative splicing" and related terms refer to all types of RNA processing that lead to expression of plural protein isoforms from a single gene. Some genes are first transcribed as long mRNA precursors that are then shortened by a series of processing steps to produce the mature mRNA molecule. One of these steps is RNA splicing, in which the intron sequences are removed from the mRNA precursor. A cell can splice

the primary transcript in different ways, making different "splice variants," and thereby making different polypeptide chains from the same gene, or from the same mRNA molecule. Splice variants can include, for example, exon insertions, exon extensions, exon truncations, exon deletions, alternatives in the 5' untranslated region and alternatives in the 3' untranslated region.

[0322]   "Oligonucleotide" may generally refer to polynucleotides of between about 5 and about 100 nucleotides of single-or double-stranded nucleic acids. For the purposes of this disclosure, there is no upper limit to the length of an oligonucleotide. Oligonucleotides are also known as oligomers or oligos and can be isolated from genes, or chemically synthesized by methods known in the art.

[0323]   "Nucleic acid composition" as used herein is a composition comprising a nucleic acid sequence, including one having an open reading frame that encodes a polypeptide and is capable, under appropriate conditions, of being expressed as a polypeptide. The term includes, for example, vectors, including plasmids, cosmids, viral vectors (e.g., retrovirus vectors such as lentivirus, adenovirus, and the like), human, yeast, bacterial, P1-derived artificial chromosomes (HAC's, YAC's, BAC's, PAC's, etc), and mini-chromosomes, *in vitro* host cells, *in vivo* host cells, tissues, organs, allogenic or congenic grafts or transplants, multicellular organisms, and chimeric, genetically modified, or transgenic animals comprising a subject nucleic acid sequence.

[0324]   An "isolated," "purified," or "substantially isolated" polynucleotide, or a polynucleotide in "substantially pure form," in "substantially purified form," in "substantial purity," or as an "isolate," is one that is substantially free of the sequences with which it is associated in nature, or other nucleic acid sequences that do not include a sequence or fragment of the subject polynucleotides. By substantially free is meant that less than about 90%, less than about 80%, less than about 70%, less than about 60%, or less than about 50% of the composition is made up of materials other than the isolated polynucleotide. For example, the isolated polynucleotide is at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97%, or at least about 99% free of the materials with which it is associated in nature. For example, an isolated polynucleotide may be present in a composition wherein at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97%, at least about 99% of the total macromolecules (for example,

polypeptides, fragments thereof, polynucleotides, fragments thereof, lipids, polysaccharides, and oligosaccharides) in the composition is the isolated polynucleotide. Where at least about 99% of the total macromolecules is the isolated polynucleotide, the polynucleotide is at least about 99% pure, and the composition comprises less than about 1% contaminant. As used herein, an "isolated," "purified" or "substantially isolated" polynucleotide, or a polynucleotide in "substantially pure form," in "substantially purified form," in "substantial purity," or as an "isolate," also refers to recombinant polynucleotides, modified, degenerate and homologous polynucleotides, and chemically synthesized polynucleotides, which, by virtue of origin or manipulation, are not associated with all or a portion of a polynucleotide with which it is associated in nature, are linked to a polynucleotide other than that to which it is linked in nature, or do not occur in nature. For example, the subject polynucleotides are generally provided as other than on an intact chromosome, and recombinant embodiments are typically flanked by one or more nucleotides not normally associated with the subject polynucleotide on a naturally-occurring chromosome.

[0325]   The terms "polypeptide," "peptide," and "protein," used interchangeably herein, refer to a polymeric form of amino acids of any length, which can include naturally-occurring amino acids, coded and non-coded amino acids, chemically or biochemically modified, derivatized, or designer amino acids, amino acid analogs, peptidomimetics, and depsipeptides, and polypeptides having modified, cyclic, bicyclic, depsicyclic, or depsibicyclic peptide backbones. The term includes single chain protein as well as multimers. The term also includes conjugated proteins, fusion proteins, including, but not limited to, GST fusion proteins, fusion proteins with a heterologous amino acid sequence, fusion proteins with heterologous and homologous leader sequences, fusion proteins with or without N-terminal methionine residues, pegolyated proteins, and immunologically tagged proteins. Also included in this term are variations of naturally occurring proteins, where such variations are homologous or substantially similar to the naturally occurring protein, as well as corresponding homologs from different species. Variants of polypeptide sequences include insertions, additions, deletions, or substitutions compared with the subject polypeptides. The term also includes peptide aptamers.

[0326]   The novel polypeptides herein include amino acid sequences encoded by an open reading frame (ORF) as shown in SEQ ID NOS.: 1232 - 2462, described

in greater detail below, including the full length protein and fragments thereof, particularly biologically active fragments and/or fragments corresponding to functional domains, e.g., a signal peptide or leader sequence, an enzyme active site, including a cleavage site and an enzyme catalytic site, a domain for interaction with other protein(s), a domain for binding DNA, a regulatory domain, a consensus domain that is shared with other members of the same protein family, such as a kinase family or an immunoglobulin family; an extracellular domain that may act as a target for antibody production or that may be cleaved to become a soluble receptor or a ligand for a receptor; an intracellular fragment of a transmembrane protein that participates in signal transduction; a transmembrane domain of a transmembrane protein that may facilitate water or ion transport; a sequence associated with cell survival and/or cell proliferation; a sequence associated with cell cycle arrest, DNA repair and/or apoptosis; a sequence associated with a disease or disease prognosis, including types of cancer, degenerative disease, inflammatory disease, immunological disease, genetic disease, metabolic disease, and/or bacterial or viral infection; and including fusions of the subject polypeptides to other proteins or parts thereof; modifications of the subject polypeptide, e.g., comprising modified, derivatized, or designer amino acids, modified peptide backbones, and/or immunological tags; as well as intra- and inter-species homologs of the subject polypeptides.

[0327] As noted above, a "biologically active" entity, or an entity having "biological activity," is one having structural, regulatory, or biochemical functions of a naturally occurring molecule or any function related to or associated with a metabolic or physiological process. Biologically active polypeptide fragments are those exhibiting activity similar, but not necessarily identical, to an activity of a polypeptide of the present invention. The biological activity can include an improved desired activity, or a decreased undesirable activity. For example, an entity demonstrates biological activity when it participates in a molecular interaction with another molecule, or when it has therapeutic value in alleviating a disease condition, or when it has prophylactic value in inducing an immune response to the molecule, or when it has diagnostic value in determining the presence of the molecule. A biologically active polypeptide or fragment thereof includes one that can participate in a biological reaction, for example, as a transcription factor that combines with other transcription factors for initiation of transcription, or that can serve as an epitope or immunogen to stimulate an immune response, such as production of antibodies, or

that can transport molecules into or out of cells, or that can perform a catalytic activity, for example polymerization or nuclease activity, or that can participate in signal transduction by binding to receptors, proteins, or nucleic acids, activating enzymes or substrates.

[0328]   A "signal peptide," or a "leader sequence," comprises a sequence of amino acid residues, typically, at the N terminus of a polypeptide, which directs the intracellular trafficking of the polypeptide. Polypeptides that contain a signal peptide or leader sequence typically also contain a signal peptide or leader sequence cleavage site. Such polypeptides, after cleavage at the cleavage sites, generate mature polypeptides, for example, after extracellular secretion or after being directed to the appropriate intracellular compartment.

[0329]   "Depsipeptides" are compounds containing a sequence of at least two alpha-amino acids and at least one alpha-hydroxy carboxylic acid, which are bound through at least one normal peptide link and ester links, derived from the hydroxy carboxylic acids. "Linear depsipeptides" can comprise rings formed through S–S bridges, or through an hydroxy or a mercapto group of an hydroxy-, or mercapto-amino acid and the carboxyl group of another amino- or hydroxy-acid but do not comprise rings formed only through peptide or ester links derived from hydroxy carboxylic acids. "Cyclic depsipeptides" are peptides containing at least one ring formed only through peptide or ester links, derived from hydroxy carboxylic acids.

[0330]   An "isolated," "purified," or "substantially isolated" polypeptide, or a polypeptide in "substantially pure form," in "substantially purified form," in "substantial purity," or as an "isolate," is one that is substantially free of the materials with which it is associated in nature or other polypeptide sequences that do not include a sequence or fragment of the subject polypeptides. By substantially free is meant that less than about 90%, less than about 80%, less than about 70%, less than about 60%, or less than about 50% of the composition is made up of materials other than the isolated polypeptide. For example, the isolated polypeptide is at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97%, or at least about 99% free of the materials with which it is associated in nature. For example, an isolated polypeptide may be present in a composition wherein at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97%, or at least about 99% of the total macromolecules (for example, polypeptides, fragments

thereof, polynucleotides, fragments thereof, lipids, polysaccharides, and oligosaccharides) in the composition is the isolated polypeptide. Where at least about 99% of the total macromolecules is the isolated polypeptide, the polypeptide is at least about 99% pure, and the composition comprises less than about 1% contaminant. As used herein, an "isolated," "purified," or "substantially isolated" polypeptide, or a polypeptide in "substantially pure form," in "substantially purified form," in "substantial purity," or as an "isolate," also refers to recombinant polypeptides, modified, tagged and fusion polypeptides, and chemically synthesized polypeptides, which by virtue or origin or manipulation, are not associated with all or a portion of the materials with which they are associated in nature, are linked to molecules other than that to which they are linked in nature, or do not occur in nature.

[0331] Detection methods of the invention can be qualitative or quantitative. Thus, as used herein, the terms "detection," "identification," "determination," and the like, refer to both qualitative and quantitative determinations, and include "measuring." For example, detection methods include methods for detecting the presence and/or level of polynucleotide or polypeptide in a biological sample, and methods for detecting the presence and/or level of biological activity of polynucleotide or polypeptide in a sample.

[0332] As used herein, the term "array" or "microarray" may be used interchangeably and refers to a collection of plural biological molecules such as nucleic acids, polypeptides, or antibodies, having locatable addresses that may be separately detectable. Generally, "microarray" encompasses use of sub microgram quantities of biological molecules. The biological molecules may be affixed to a substrate or may be in solution or suspension. The substrate can be porous or solid, planar or non-planar, unitary or distributed, such as a glass slide, a 96 well plate, with or without the use of microbeads or nanobeads. As such, the term "microarray" includes all of the devices referred to as microarrays in Schena, 1999; Bassett et al., 1999; Bowtell, 1999; Brown and Botstein, 1999; Chakravarti, 1999; Cheung et al., 1999; Cole et al., 1999; Collins, 1999; Debouck and Goodfellow, 1999; Duggan et al., 1999; Hacia, 1999; Lander, 1999; Lipshutz et al., 1999; Southern, et al., 1999; Schena, 2000; Brenner et al, 2000; Lander, 2001; Steinhaur et al., 2002; and Espejo et al, 2002. Nucleic acid microarrays include both oligonucleotide arrays (DNA chips) containing expressed sequence tags ("ESTs") and arrays of larger DNA sequences representing a plurality of genes bound to the substrate, either one of which can be

used for hybridization studies. Protein and antibody microarrays include arrays of polypeptides or proteins, including but not limited to, polypeptides or proteins obtained by purification, fusion proteins, and antibodies, and can be used for specific binding studies (Zhu and Snyder, 2003; Houseman et al., 2002; Schaeferling et al., 2002; Weng et al., 2002; Winssinger et al., 2002; Zhu et al., 2001; Zhu et al. 2001; and MacBeath and Schreiber, 2000).

[0333] A "nucleic acid hybridization reaction" is one in which single strands of DNA or RNA randomly collide with one another, and bind to each other only when their nucleotide sequences have some degree of complementarity. The solvent and temperature conditions can be varied in the reactions to modulate the extent to which the molecules can bind to one another. Hybridization reactions can be performed under different conditions of "stringency." The "stringency" of a hybridization reaction as used herein refers to the conditions (e.g., solvent and temperature conditions) under which two nucleic acid strands will either pair or fail to pair to form a "hybrid" helix.

[0334] "T_m" is the temperature in degrees Celsius at which 50% of a polynucleotide duplex made of complementary strands of nucleic acids that are hydrogen bonded in an anti-parallel direction by Watson-Crick base pairing dissociate into single strands under conditions of the hybridization reaction. T_m can be predicted according to a standard formula, such as: $T_m = 81.5 + 16.6 \log[X^+] + 0.41 (\%G/C) - 0.61 (\%F) - 600/L$, where $[X^+]$ is the cation concentration (usually sodium ion, Na^+) in mol/L; (%G/C) is the number of G and C residues as a percentage of total residues in the duplex; (%F) is the percent formamide in solution (wt/vol); and L is the number of nucleotides in each strand of the paired nucleic acids.

[0335] A "buffer" is a system that tends to resist change in pH when a given increment of hydrogen ion or hydroxide ion is added. Buffered solutions contain conjugate acid-base pairs. Any conventional buffer can be used with the inventions herein including but not limited to, for example, Tris, phosphate, imidazole, and bicarbonate.

[0336] A "library" of polynucleotides comprises a collection of sequence information of a plurality of polynucleotide sequences, which information is provided in either biochemical form (e.g., as a collection of polynucleotide molecules), or in electronic form (e.g., as a collection of polynucleotide sequences stored in a

computer-readable form, as in a computer-based system, a computer data file, and/or as part of a computer program).

[0337] A "library" of polypeptides comprises a collection of sequence information of a plurality of polypeptide sequences, which information is provided in, e.g., a collection of polypeptide sequences stored in a computer-readable form, as in a computer-based system, a computer data file, and/or as part of a computer program.

[0338] "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the genome sequence or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid, e.g., with computer-readable media comprising data storage structures. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

[0339] "Recorded" refers to a process for storing information on computer readable media, using any such methods as known in the art.

[0340] As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

[0341] "Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif, or expression levels of a polynucleotide in a sample, with the stored sequence information. A variety of known algorithms are publicly known and commercially available, e.g., MacPattern (EMBL), BLAST, BLASTN and BLASTX (NCBI), gapped BLAST, BLAZE, the Wise package, FASTX, Clustalw, FASTA, FASTA3, Align0, TCoffee, BestFit, FastDB, and TeraBLAST (TimeLogic, Crystal Bay, Nevada). Search means can be used to identify fragments or regions of the genome

100

that match a particular target sequence or target motif, for example, based on sequence similarity, for example, to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

[0342] "Sequence similarity," "sequence homology," "homology," "sequence identity," and "percent sequence identity," used interchangeably herein, describe the degree of relatedness between two polynucleotide or polypeptide sequences. In general, "identity" means the exact match-up of two or more nucleotide sequences or two or more amino acid sequences, where the nucleotide or amino acids being compared are the same. Also, in general, "similarity" or "homology" means the exact match-up of two or more nucleotide sequences or two or more amino acid sequences, where the nucleotide or amino acids being compared are either the same or possess similar chemical and/or physical properties. The terms also refer to the percentage of the "aligned" bases (for the polynucleotides) or amino acid residues (for the polypeptides) that are identical when the sequences are aligned. Sequences can be aligned in a number of different ways and sequence similarity can be determined in a number of different ways. For example, the bases or amino acid residues of one sequence can be aligned to a gap in the other sequence, or they can be aligned only to another base or amino acid residue in the other sequence. A gap can range anywhere from one nucleotide, base, or amino acid residue to multiple exons in length, up to any number of nucleotides or amino acid residues. Further, sequences can be aligned such that nucleotides (or bases) align with nucleotides, nucleotides align with amino acid residues, or amino acid residues align with amino acid residues.

[0343] A "target sequence" can be any polynucleotide or amino acid sequence of six or more contiguous nucleotides or two or more amino acids, for example, from about 5 or from about 10 to about 100 amino acids, or from about 15 or from about 30 to about 300 nucleotides. A variety of comparing means can be used to accomplish comparison of sequence information from a sample (e.g., to analyze target sequences, target motifs, or relative expression levels) with the data storage means. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention to accomplish comparison of target sequences and motifs. Computer programs to analyze expression levels in a sample and in controls are also known in the art. A "target sequence" includes an "antibody target sequence," which refers to an amino acid sequence that can be used as an

immunogen for injection into animals for production of antibodies or for screening against a phage display or antibody library for identification of binding partners.

[0344] A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences, and other expression elements such as binding sites for transcription factors.

[0345] A "matrix" is a geometric network of antibody molecules and their antigens, as found in immunoprecipitation and flocculation reactions. An antibody matrix can exist in solution or on a solid phase support.

[0346] The term "binds specifically," in the context of antibody binding, refers to high avidity and/or high affinity binding of an antibody to a specific polypeptide, or more accurately, to an epitope of a specific polypeptide. Antibody binding to such epitope on a polypeptide can be stronger than binding of the same antibody to any other epitopes, particularly other epitopes that can be present in molecules in association with, or in the same sample as the polypeptide of interest. For example, when an antibody binds more strongly to one epitope than to another, adjusting the binding conditions can result in antibody binding almost exclusively to the specific epitope and not to any other epitopes on the same polypeptide, and not to any other polypeptide, which does not comprise the epitope. Antibodies that bind specifically to a subject polypeptide may be capable of binding other polypeptides at a weak, yet detectable, level (e.g., 10% or less of the binding shown to the polypeptide of interest). Such weak binding, or background binding, is readily discernible from the specific antibody binding to a subject polypeptide, e.g., by use of appropriate controls. In general, antibodies of the invention bind to a specific polypeptide with a binding affinity of 10^{-7} M or greater (e.g., 10^{-8} M, 10^{-9} M, 10^{-10}, 10^{-11}, etc.).

[0347] The term "host cell" includes an individual cell, cell line, cell culture, or *in vivo* cell, which can be or has been a recipient of any polynucleotides or polypeptides of the invention, for example, a recombinant vector, an isolated polynucleotide, antibody or fusion protein. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in

morphology, physiology, or in total DNA, RNA, or polypeptide complement) to the original parent cell due to natural, accidental, or deliberate mutation and/or change. Host cells can be prokaryotic or eukaryotic, including mammalian, insect, amphibian, reptile, crustacean, avian, fish, plant and fungal cells. A host cell includes cells transformed, transfected, transduced, or infected *in vivo* or *in vitro* with a polynucleotide of the invention, for example, a recombinant vector. A host cell which comprises a recombinant vector of the invention may be called a "recombinant host cell."

[0348]   "Biological sample," "patient sample," "clinical sample" "sample," or "biological specimen," used interchangeably herein, encompasses a variety of sample types obtained from an individual, including biological fluids such as blood, serum, plasma, urine, cerebrospinal fluid, tears, saliva, lymph, dialysis fluid, lavage fluid, semen, and other liquid samples or tissues of biological origin. It includes tissue samples and tissue cultures or cells derived therefrom and the progeny thereof, including cells in culture, cell supernatants, and cell lysates. It includes organ or tissue culture derived fluids, tissue biopsy samples, tumor biopsy samples, stool samples, and fluids extracted from physiological tissues. Cells dissociated from solid tissues, tissue sections, and cell lysates are included. The definition also includes samples that have been manipulated in any way after their procurement, such as by treatment with reagents, solubilization, or enrichment for certain components, such as polynucleotides or polypeptides. Also included in the term are derivatives and fractions of biological samples. A biological sample can be used in a diagnostic, monitoring, or screening assay.

[0349]   The terms "individual," "host," "patient," and "subject," used interchangeably herein, refer to a mammal, including, but not limited to, murines, simians, humans, felines, canines, equines, bovines, porcines, ovines, caprines, mammalian farm animals, mammalian sport animals, and mammalian pets. "Mammals" or "mammalian," are used broadly to describe organisms which are within the class mammalia, including the orders carnivore (*e.g.*, dogs and cats), rodentia (*e.g.*, mice, guinea pigs, and rats), and other mammals, including cattle, goats, sheep, cows, horses, rabbits, and pigs, and primates (*e.g.*, humans, chimpanzees, and monkeys).

[0350]   The terms "agent," "substance," "modulator," and "compound" are used interchangeably herein. These terms refer to a substance that binds to or

103

modulates a level or activity of a subject polypeptide or a level of mRNA encoding a subject protein or nucleic acid, or that modulates the activity of a cell containing the subject protein or nucleic acid . Where the agent modulates a level of mRNA encoding a subject protein, agents include ribozymes, antisense, and RNAi molecules. Where the agent is a substance that modulates a level of activity of a subject polypeptide, agents include antibodies specific for the subject polypeptide, peptide aptamers, small molecules, agents that bind a ligand-binding site in a subject polypeptide, and the like. Antibody agents include antibodies that specifically bind a subject polypeptide and activate the polypeptide, such as receptor-ligand binding that initiates signal transduction; antibodies that specifically bind a subject polypeptide and inhibit binding of another molecule to the polypeptide, thus preventing activation of a signal transduction pathway; antibodies that bind a subject polypeptide to modulate transcription; antibodies that bind a subject polypeptide to modulate translation; as well as antibodies that bind a subject polypeptide on the surface of a cell to initiate antibody-dependent cytotoxicity ("ADCC") or to initiate cell killing or cell growth. Small molecule agents include those that bind the polypeptide to modulate activity of the polypeptide or cell containing the polypeptide in a similar fashion. The term "agent" also refers to substances that modulate a condition or disorder associated with a subject polynucleotide or polypeptide. Such agents include subject polynucleotides themselves, subject polypeptides themselves, and the like. Agents may be chosen from amongst candidate agents, as defined below.

[0351] The terms "candidate agent," "subject agent," or "test agent," used interchangeably herein, encompass numerous chemical classes, typically synthetic, semi-synthetic, or naturally occurring inorganic or organic molecules, small molecules, or macromolecular complexes. Candidate agents can be small organic compounds having a molecular weight of more than about 50 and less than about 2,500 daltons. Candidate agents can comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and can include at least an amine, carbonyl, hydroxyl or carboxyl group, and can contain at least two of the functional chemical groups. The candidate agents can comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules, including oligonucleotides, polynucleotides, and fragments thereof, depsipeptides, polypeptides and fragments thereof, oligosaccharides, polysaccharides

104

and fragments thereof, lipids, fatty acids, steroids, purines, pyrimidines, derivatives thereof, structural analogs, modified nucleic acids, modified, derivatized or designer amino acids, or combinations thereof.

[0352] An "agent which modulates a biological activity of a subject polypeptide," as used herein, describes any substance, synthetic, semi-synthetic, or natural, organic or inorganic, small molecule or macromolecular, pharmaceutical or protein, with the capability of altering a biological activity of a subject polypeptide or of a fragment thereof, as described herein. Generally, a plurality of assay mixtures is run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e., at zero concentration or below the level of detection. The biological activity can be measured using any assay known in the art.

[0353] An agent which modulates a biological activity of a subject polypeptide increases or decreases the activity at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 50%, at least about 100%, or at least about 2-fold, at least about 5-fold, or at least about 10-fold or more when compared to a suitable control.

[0354] The term "agonist" refers to a substance that mimics the function of an active molecule. Agonists include, but are not limited to, drugs, hormones, antibodies, and neurotransmitters, as well as analogues and fragments thereof.

[0355] The term "antagonist" refers to a molecule that competes for the binding sites of an agonist, but does not induce an active response. Antagonists include, but are not limited to, drugs, hormones, antibodies, and neurotransmitters, as well as analogues and fragments thereof.

[0356] The term "receptor" refers to a polypeptide that binds to a specific extracellular molecule and may initiate a cellular response.

[0357] The term "ligand" refers to any molecule that binds to a specific site on another molecule.

[0358] The term "modulate" encompasses an increase or a decrease, a stimulation, inhibition, or blockage in the measured activity when compared to a suitable control. "Modulation" of expression levels includes increasing the level and decreasing the level of an mRNA or polypeptide encoded by a polynucleotide of the invention when compared to a control lacking the agent being tested. In some embodiments, agents of particular interest are those which inhibit a biological activity

of a subject polypeptide, and/or which reduce a level of a subject polypeptide in a cell, and/or which reduce a level of a subject mRNA in a cell and/or which reduce the release of a subject polypeptide from a eukaryotic cell. In other embodiments, agents of interest are those that increase a biological activity of a subject polypeptide, and/or which increase a level of a subject polypeptide in a cell, and/or which increase a level of a subject mRNA in a cell and/or which increase the release of a subject polypeptide from a eukaryotic cell.

[0359]   An agent that "modulates the level of expression of a nucleic acid" in a cell is one that brings about an increase or decrease of at least about 1.25-fold, at least about 1.5-fold, at least about 2-fold, at least about 5-fold, at least about 10-fold, or more in the level (i.e., an amount) of mRNA and/or polypeptide following cell contact with a candidate agent compared to a control lacking the agent.

[0360]   "Modulating a level of active subject polypeptide" includes increasing or decreasing activity of a subject polypeptide; increasing or decreasing a level of active polypeptide protein; increasing or decreasing a level of mRNA encoding active subject polypeptide, and increasing or decreasing the release of subject polypeptide for a eukaryotic cell. In some embodiments, an agent is a subject polypeptide, where the subject polypeptide itself is administered to an individual. In some embodiments, an agent is an antibody specific for a subject polypeptide. In some embodiments, an agent is a chemical compound such as a small molecule that may be useful as an orally available drug. Such modulation includes the recruitment of other molecules that directly effect the modulation. For example, an antibody that modulates the activity of a subject polypeptide that is a receptor on a cell surface may bind to the receptor and fix complement, activating the complement cascade and resulting in lysis of the cell.

[0361]   The term "over-expressed" refers to a state wherein there exists any measurable increase over normal or baseline levels. For example, a molecule that is over-expressed in a disorder is one that is manifest in a measurably higher level compared to levels in the absence of the disorder.

[0362]   "Treatment," "treating," and the like, as used herein, refer to obtaining a desired pharmacologic and/or physiologic effect, covering any treatment of a pathological condition or disorder in a mammal, including a human. The effect may be prophylactic in terms of completely or partially preventing a disorder or symptom thereof and/or may be therapeutic in terms of a partial or complete cure for

a disorder and/or adverse affect attributable to the disorder. That is, "treatment" includes (1) preventing the disorder from occurring or recurring in a subject who may be predisposed to the disorder but has not yet been diagnosed as having it, (2) inhibiting the disorder, such as arresting its development, (3) stopping or terminating the disorder or at least symptoms associated therewith, so that the host no longer suffers from the disorder or its symptoms, such as causing regression of the disorder or its symptoms, for example, by restoring or repairing a lost, missing or defective function, or stimulating an inefficient process, or (4) relieving, alleviating, or ameliorating the disorder, or symptoms associated therewith, where ameliorating is used in a broad sense to refer to at least a reduction in the magnitude of a parameter, such as inflammation, pain, and/or tumor size.

[0363] A "pharmaceutically acceptable carrier," "pharmaceutically acceptable diluent," or "pharmaceutically acceptable excipient," or "pharmaceutically acceptable vehicle," used interchangeably herein, refer to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any conventional type. A pharmaceutically acceptable carrier is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the carrier for a formulation containing polypeptides would not normally include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Suitable carriers include, but are not limited to, water, dextrose, glycerol, saline, ethanol, and combinations thereof. The carrier can contain additional agents such as wetting or emulsifying agents, pH buffering agents, or adjuvants which enhance the effectiveness of the formulation. Adjuvants of the invention include, but are not limited to Freunds's, Montanide ISA Adjuvants [Seppic, Paris, France], Ribi's Adjuvants (Ribi ImmunoChem Research, Inc., Hamilton, MT), Hunter's TiterMax (CytRx Corp., Norcross, GA), Aluminum Salt Adjuvants (Alhydrogel - Superfos of Denmark/Accurate Chemical and Scientific Co., Westbury, NY), Nitrocellulose-Adsorbed Protein, Encapsulated Antigens, and Gerbu Adjuvant (Gerbu Biotechnik GmbH, Gaiberg, Germany/C-C Biotech, Poway, CA). Topical carriers include liquid petroleum, isopropyl palmitate, polyethylene glycol, ethanol (95%), polyoxyethylene monolaurate (5%) in water, or sodium lauryl sulfate (5%) in water. Other materials such as anti-oxidants, humectants, viscosity stabilizers, and similar agents can be added as necessary. Percutaneous penetration enhancers such as Azone can also be included.

[0364] "Pharmaceutically acceptable salts" include the acid addition salts (formed with the free amino groups of the polypeptide) and which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, mandelic, oxalic, and tartaric. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, and histidine.

[0365] Compositions for oral administration can form solutions, suspensions, tablets, pills, capsules, sustained release formulations, oral rinses, or powders.

[0366] The term "unit dosage form," as used herein, refers to physically discrete units suitable as unitary dosages for human and animal subjects, each unit containing a predetermined quantity of compounds of the present invention calculated in an "effective amount," that is, a dosage sufficient to produce the desired result or effect in association with a pharmaceutically acceptable carrier. The specifications for the novel unit dosage forms of the present invention depend on the particular compound employed, the host, and the effect to be achieved, as well as the pharmacodynamics associated with each compound in the host.

Compositions

[0367] The present invention provides novel isolated polynucleotides encoding polypeptides and fragments thereof. The present invention also provides novel isolated polypeptides, fragments thereof, and compositions comprising same. The present invention further provides polynucleotide compositions that can be used to identify the polypeptides.

[0368] The present invention provides recombinant vectors and host cells for use in gene expression, primer pairs for use in hybridizations, computer-based embodiments for use in bioinformatics, and transgenic animals and embryonic stem cell lines for use in mutating and regulating gene expression.

Nucleic Acids

<u>Sequences</u>

[0369] This invention provides genes encoding proteins, the encoded proteins, and fragments and homologs thereof. It provides human polynucleotide sequences and the corresponding mouse polynucleotide sequences.

[0370]   The nucleic acids of the subject invention can encode all or a part of the subject proteins. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, for example by restriction enzyme digestion or polymerase chain reaction (PCR) amplification. The use of the polymerase chain reaction has been described (Saiki et al., 1985) and current techniques have been reviewed (Sambrook et al., 1989; McPherson et al. 2000; Dieffenbach and Dveksler, 1995). For the most part, DNA fragments will be of at least about 5 nucleotides, at least about 8 nucleotides, at least about 10 nucleotides, at least about 15 nucleotides, at least about 18 nucleotides, at least about 20 nucleotides, at least about 25 nucleotides, at least about 30 nucleotides, or at least about 50 nucleotides, at least about 75 nucleotides, or at least about 100 nucleotides. Nucleic acid compositions that encode at least six contiguous amino acids (i.e., fragments of 18 nucleotides or more), for example, nucleic acid compositions encoding at least 8 contiguous amino acids (i.e., fragments of 24 nucleotides or more), are useful in directing the expression or the synthesis of peptides that can be used as immunogens (Lerner, 1982; Shinnick et al., 1983; Sutcliffe et al., 1983).

[0371]   In some embodiments, a polynucleotide of the invention comprises a nucleotide sequence of at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, at least about 900, at least about 950, at least about 1000, at least about 1100, at least about 1200, at least about 1300, at least about 1400, at least about 1500, at least about 1600, at least about 1700, at least about 1800, at least about 1900, at least about 2000, at least about 2100, at least about 2200, at least about 2300, at least about 2400, at least about 2500, at least about 3000, at least about 4000, or at least about 5000 contiguous nucleotides of any one of the sequences shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697, or the coding region thereof, or a complement thereof.

[0372]   In other embodiments, a polynucleotide of the invention has at least about 60%, 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 97%, at least about 98%, or at least

about 99% nucleotide sequence identity with a nucleotide sequence, or a fragment thereof, of the coding region of any one of the sequences shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697, or a complement thereof. These sequence variants include naturally-occurring variants (e.g., SNPs, allelic variants, and homologs from other species), degenerate variants, variants associated with disease or pathological states, and variants resulting from random or directed mutagenesis, as well as from chemical or other modification.

[0373] In some embodiments, a polynucleotide of the invention comprises a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence of at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, at least about 900, at least about 950, or at least about 1000 contiguous amino acids of at least one of the sequences shown in SEQ ID NOS.: 1232 - 2462 (e.g., a polypeptide encoded by at least one of the nucleotide sequences shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697), up to and including an entire amino acid sequence as shown in SEQ ID NOS.: 1232 - 2462 (or as encoded by at least one of the nucleotide sequences shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697).

[0374] In some embodiment, the present invention includes the present polynucleotide selected from SEQ ID NOS.: 1 - 1231 and 2463 - 3697, which contain 300 bp of 5' terminus of a protein encoding polynucleotide sequence. Such a polynucleotide is useful for the purposes of clustering gene sequences to determine gene family.

[0375] In further embodiments, a polynucleotide of the invention hybridizes under stringent hybridization conditions to a polynucleotide having the coding region of any one of the sequences shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697, or a complement thereof.

[0376] The polynucleotides of the invention include those that encode variants of the polypeptide sequences encoded by the polynucleotides of the Sequence Listing. In some embodiments, these polynucleotides encode variant polypeptides that include insertions, additions, deletions, or substitutions compared with the

polypeptides encoded by the nucleotide sequences shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697, and in Table 1.. Conservative amino acid substitutions include serine/threonine, valine/leucine/isoleucine, asparagine/histidine/glutamine, glutamic acid/aspartic acid, etc. (Gonnet et al., 1992).

[0377] The nucleic acids of the invention include degenerate variants that can be translated, according to the standard genetic code, to provide an amino acid sequence identical to that translated from the nucleic acid sequences herein. For example, synonymous codons include GGG, GGA, GGC, and GGU, each encoding Glycine.

[0378] The nucleic acids of the invention include single nucleotide polymorphisms (SNPs), which occur frequently in eukaryotic genomes (Lander, et al. 2001). The nucleotide sequence determined from one individual of a species can differ from other allelic forms present within the population.

[0379] The nucleic acids of the invention include homologs of the polynucleotides. The source of homologous genes can be any species, e.g., primate species, particularly human; rodents, such as rats, hamsters, guinea pigs, and mice; rabbits, canines, felines; cattles, such as bovines, goats, pigs, sheep, equines, crustaceans, birds, chickens, reptiles, amphibians, fish, insects, plants, fungi, yeast, nematodes, etc. Among mammalian species, e.g., human and mouse, homologs have substantial sequence similarity, e.g., at least about 60% sequence identity, at least about 75% sequence identity, or at least about 80% sequence identity among nucleotide sequences. In many embodiments of interest, homology will be at least about 75%, at least about 80% ,at least about 85%, at least about 90%, at least about 95%, at least about 97%, or at least about 98%, where in certain embodiments of interest homology will be as high as about 99%.

[0380] Modifications in the native structure of nucleic acids, including alterations in the backbone, sugars or heterocyclic bases, have been shown to increase intracellular stability and binding affinity. Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O- phosphorothioate, 3'-CH_2-5'-O-phosphonate and 3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire ribose phosphodiester backbone with a peptide linkage.

111

[0381] Sugar modifications are also used to enhance stability and affinity. The α-anomer of deoxyribose can be used, where the base is inverted with respect to the natural β-anomer. The 2'-OH of the ribose sugar can be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity.

[0382] Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'- deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5-propynyl-2'- deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

[0383] A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, about 2 kb, and possibly more, of flanking genomic DNA at either the 5' or 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' or 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue and stage specific expression.

[0384] Nucleic acid molecules of the invention can comprise heterologous nucleic acid molecules, i.e., nucleic acid molecules other than the subject nucleic acid molecules, of any length. For example, the subject nucleic acid molecules can be flanked on the 5' and/or 3' ends by heterologous nucleic acid molecules of from about 1 nucleotide to about 10 nucleotides, from about 10 nucleotides to about 20 nucleotides, from about 20 nucleotides to about 50 nucleotides, from about 50 nucleotides to about 100 nucleotides, from about 100 nucleotides to about 250 nucleotides, from about 250 nucleotides to about 500 nucleotides, or from about 500 nucleotides to about 1000 nucleotides, or more in length.

[0385] The subject polynucleotides include those that encode fusion proteins comprising the subject polypeptides fused to "fusion partners." For example, the present soluble receptor or ligand can be fused to an immunoglobulin fragment, such

as an Fc fragment for stability in circulation or to fix complement. Other polypeptide fragments that have equivalent capabilities as the Fc fragments can also be used herein.

[0386] The isolated nucleic acids of the invention can be used as probes to detect and characterize gross alteration in a genomic locus, such as deletions, insertions, translocations, and duplications, e.g., applying fluorescence *in situ* hybridization (FISH) techniques to examine chromosome spreads (Andreeff et al., 1999). The nucleic acids are also useful for detecting smaller genomic alterations, such as deletions, insertions, additions, translocations, and substitutions (e.g., SNPs).

[0387] When used as probes to detect nucleic acid molecules capable of hybridizing with nucleic acids described in the Sequence Listing, the nucleic acid molecules can be flanked by heterologous sequences of any length. When used as probes, a subject nucleic acid can include nucleotide analogs that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogs that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin, digoxigenin, and dinitrophenyl.

[0388] Suitable fluorescent labels include fluorochromes e.g., fluorescein and its derivatives, e.g., fluorescein isothiocyanate (FITC6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE),), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM); coumarin and its derivatives, e.g., 7-amino-4-methylcoumarin, aminocoumarin; bodipy dyes, such as Bodipy FL; cascade blue; Oregon green; rhodamine dyes, e.g., rhodamine, 6-carboxy-X-rhodamine (ROX), Texas red, phycoerythrin, and tetramethylrhodamine; eosins and erythrosins; cyanine dyes, e.g., allophycocyanin, Cy3 and Cy5 or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); macrocyclic chelates of lanthanide ions, e.g., quantum dye, etc; and chemiluminescent molecules, e.g., luciferases.

[0389] Fluorescent labels also include a green fluorescent protein (GFP), i.e., a "humanized" version of a GFP, e.g., wherein codons of the naturally-occurring nucleotide sequence are changed to more closely match human codon bias; a GFP derived from *Aequoria victoria* or a derivative thereof, e.g., a "humanized" derivative such as Enhanced GFP, which are available commercially, e.g., from Clontech, Inc.; other fluorescent mutants of a GFP from *Aequoria victoria*, e.g., as described in U.S.

Patent No. 6,066,476; 6,020,192; 5,985,577; 5,976,796; 5,968,750; 5,968,738; 5,958,713; 5,919,445; 5,874,304; a GFP from another species such as *Renilla reniformis, Renilla mulleri,* or *Ptilosarcus guernyi,* as previously described (WO 99/49019; Peelle et al., 2001), "humanized" recombinant GFP (hrGFP) (Stratagene®); any of a variety of fluorescent and colored proteins from Anthozoan species, (e.g., Matz et al., 1999).

[0390]   Probes can also contain fluorescent analogs, including commercially available fluorescent nucleotide analogs that can readily be incorporated into a subject nucleic acid. These include deoxyribonucleotides and/or ribonucleotide analogs labeled with Cy3, Cy5, Texas Red, Alexa Fluor dyes, rhodamine, cascade blue, or BODIPY, and the like.

[0391]   Suitable radioactive labels include, e.g., ^{32}P, ^{35}S, or ^{3}H. For example, probes can contain radiolabeled analogs, including those commonly labeled with ^{32}P or ^{35}S, such as α-^{32}P-dATP, -dTTP, -dCTP, and dGTP; γ-^{35}S-GTP and α-^{35}S-dATP, and the like.

[0392]   Nucleic acids of the invention can also be bound to a substrate. Subject nucleic acids can be attached covalently, attached to a surface of the support or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence, e.g., by noncovalent interactions, or some combination thereof. The nucleic acids can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of the bound nucleic acids being separately detectable.

[0393]   The substrate can be porous or solid, planar or non-planar, unitary or distributed; and the bond between the nucleic acid and the substrate can be covalent or non-covalent. The substrate can be in the form of microbeads or nanobeads. Substrates include, but are not limited to, a membrane, such as nitrocellulose, nylon, positively-charged derivatized nylon; a solid substrate such as glass, amorphous silicon, crystalline silicon, plastics (including e.g., polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, cellulose acetate, or mixtures thereof).

[0394]   The subject nucleic acids include antisense RNA, ribozymes, and RNAi. Further, The nucleic acids of the invention can be used for antisense or RNAi inhibition of transcription or translation using methods known in the art (Phillips,

114

1999a; Phillips, 1999b; Hartmann et al., 1999; Stein et al., 1998; Agrawal et al., 1998).

Expression Vectors

[0395]　The instant invention further provides host cells, e.g., recombinant host cells, that comprise a subject nucleic acid, host cells that comprise a recombinant vector, and host cells that secrete antibodies of the invention. Subject host cells can be cultured *in vitro*, or can be part of a multicellular organism. Host cells are described in more detail below. The instant invention further provides transgenic plants and non-human animals, as described in more detail below.

[0396]　In addition to the plurality of uses described in greater detail in following sections, the subject nucleic acids find use in the preparation of all or a portion of the polypeptides of the subject invention, as described above, using an expression system. For expression, an expression vector can be employed. The expression vector will provide a transcriptional and translational initiation region, which may be inducible, conditionally-active, or constitutive, or tissue-specific, where the coding region is operably linked under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational termination region. These control regions can be native to a gene encoding the subject peptides, or can be derived from heterologous or exogenous sources.

[0397]　The subject nucleic acids can also be provided as part of a vector (e.g., a polynucleotide construct comprising an expression cassette), a wide variety of which are known in the art. Vectors include, but are not limited to, plasmids; cosmids; viral vectors; human, yeast, bacterial, P1-derived artificial chromosomes (HAC's, YAC's, BAC's, PAC's, etc.), mini-chromosomes, and the like. Vectors are amply described in numerous publications well known to those in the art (Ausubel, et al.; Jones et al., 1998a; Jones et al., 1998b). Vectors can provide for nucleic acid expression, for nucleic acid propagation, or both.

[0398]　A recombinant vector or construct that includes a nucleic acid of the invention is useful for propagating a nucleic acid in a host cell; such vectors are known as "cloning vectors." Vectors can transfer nucleic acid between host cells derived from disparate organisms; these are known in the art as "shuttle vectors." Vectors can also insert a subject nucleic acid into a host cell's chromosome; these are known in the art as "insertion vectors." Vectors can express either sense or antisense RNA transcripts of the invention *in vitro* (e.g., in a cell-free system or within an *in*

vitro cultured host cell) or *in vivo* (e.g., in a multicellular plant or animal); these are known in the art as "expression vectors," which can be part of an expression system. Expression vectors can also produce a subject antibody.

[0399] Vectors typically include at least one origin of replication, at least one site for insertion of heterologous nucleic acid (e.g., in the form of a polylinker with multiple, tightly clustered, single cutting restriction endonuclease recognition sites), and at least one selectable marker, although some integrative vectors will lack an origin that is functional in the host to be chromosomally modified, and some vectors will lack selectable markers. Vectors are transiently or stably be maintained in the cells, usually for a period of at least about one day, at least about several days to at least about several weeks.

[0400] Promoters of the invention can be naturally contiguous or not naturally contiguous to the expressed nucleic acid molecule. The promoters can be inducible, conditionally active (such as the cre-lox promoter), constitutive, and/or tissue specific.

[0401] Prior to vector insertion, the DNA of interest will be obtained substantially free of other nucleic acid sequences. The DNA can be "recombinant," and flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

[0402] Expression vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences encoding heterologous protein or RNA molecules. A selectable marker operative in the expression system or host can be present. Expression vectors can be used for the production of fusion proteins, where the fusion peptide provides additional functionality, i.e., increased protein synthesis, a leader sequence for secretion, stability, reactivity with defined antisera, or an enzyme marker, e.g., β-galactosidase.

[0403] Expression vectors can be prepared comprising a transcription cassette comprising a transcription initiation region, the gene or fragment thereof, and a transcriptional termination region. Of particular interest is the use of DNA sequences that allow for the expression of functional epitopes or domains, at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about

300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, at least about 900, at least about 950, or at least about 1000 amino acids in length, or any of the above-described fragments, up to and including the complete open reading frame of the gene. After introduction of these DNA sequences, the cells containing the vector construct can be selected by means of a selectable marker, and the selected cells expanded and used as expression-competent host cells.

[0404]   Host cells can comprise prokaryotes or eukaryotes that express proteins and polypeptides in accordance with conventional methods, the method depending on the purpose for expression. For large scale production of the protein, a unicellular organism, such as *E. coli, B. subtilis, S. cerevisiae*, insect cells in combination with baculovirus vectors, or cells of a higher organism such as vertebrates, particularly mammals, e.g., COS 7 cells, can be used as the expression host cells. In some situations, it is desirable to express eukaryotic genes in eukaryotic cells, where the encoded protein will benefit from native folding and post-translational modifications.

[0405]   Specific expression systems of interest include plants, bacteria, yeast, insect cells, and mammalian cell-derived expression systems. Representative systems from each of these categories are provided below.

[0406]   Expression systems in plants include those described in U.S. Patent No. 6,096,546 and U.S. Patent No. 6,127,145.

[0407]   Expression systems in bacteria include those described by Chang et al., 1978; Goeddel et al., 1979; Goeddel et al., 1980; EP 0 036,776; U.S. Patent No. 4,551,433; DeBoer et al., 1983); and Siebenlist et al., 1980.

[0408]   Expression systems in yeast include those described by Hinnen et al., 1978; Ito et al., 1983; Kurtz et al., 1986; Kunze et al., 1985; Gleeson et al., 1986; Roggenkamp et al., 1986; Das et al., 1984; De Louvencourt et al., 1983; Van den Berg et al., 1990; Kunze et al., 1985; Cregg et al., 1985; U.S. Patent Nos. 4,837,148 and 4,929,555; Beach and Nurse, 1981; Davidow et al., 1985; Gaillardin et al., 1985; Ballance et al., 1983; Tilburn et al., 1983; Yelton et al., 1984; Kelly and Hynes, 1985; EP 0 244,234; WO 91/00357; and U.S. Patent No. 6,080,559.

[0409]   Expression systems for heterologous genes in insects include those described in U.S. Patent No. 4,745,051; Friesen et al., 1986; EP 0 127,839; EP 0

155,476; Vlak et al., 1988; Miller et al., 1988; Carbonell et al., 1988; Maeda et al., 1985; Lebacq-Verheyden et al., 1988; Smith et al., 1985); Miyajima et al., 1987; and Martin et al., 1988. Numerous baculoviral strains and variants and corresponding permissive insect host cells are described in Luckow et al., 1988, Miller et al., 1986, and Maeda et al., 1985. The insect cell expression system is useful not only for production of heterologous proteins intracellularly, but can be used for expression of transmembrane proteins on the insect cell surfaces. Such insect cells can be used as immunogen for production of antibodies, for example, by injection of the insect cells into mice or rabbits or other suitable animals, for production of antibodies.

[0410]   Mammalian expression systems include those described in Dijkema et al., 1985; Gorman et al., 1982; Boshart et al., 1985; and U.S. Patent No. 4,399,216. Additional features of mammalian expression are facilitated as described in Ham and Wallace, 1979; Barnes and Sato, 1980 U.S. Patent Nos. 4,767,704, 4,657,866, 4,927,762, 4,560,655, WO 90/103430, WO 87/00195, and U.S. RE 30,985. Mammalian cell expression systems can also be used for production of antibodies.

[0411]   The present polynucleotides can also be used in cell-free expression systems such as bacterial system, e.g., E. coli lysate, rabbit reticulocyte lysate system, wheat germ extract system, frog oocyte lysate system, and the like which is conventional in the art. See, for example, WO 00/68412, WO 01/27260, WO 02/24939, WO 02/38790, WO 91/02076, and WO 91/02075.

[0412]   When any of the above-referenced host cells, or other appropriate host cells or organisms, are used to replicate and/or express the polynucleotides of the invention, the resulting replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism.

[0413]   Once the gene corresponding to a selected polynucleotide is identified, its expression can be regulated in the gene's native cell types. For example, an endogenous gene of a cell can be regulated by an exogenous regulatory sequence inserted into the genome of the cell at a location that will enhance or reduce expression of the gene corresponding to the subject polypeptide. The regulatory sequence can be designed to integrate into the genome via homologous recombination, as disclosed in U.S. Patent Nos. 5,641,670 and 5,733,761, the disclosures of which are herein incorporated by reference. Alternatively, it can be designed to integrate into the genome via non-homologous recombination, as

described in WO 99/15650, the disclosure of which is also herein incorporated by reference. Also encompassed in the subject invention is the production of proteins without manipulating the encoding nucleic acid itself, but rather by integrating a regulatory sequence into the genome of a cell that already includes a gene that encodes the protein of interest; this production method is described in the above-incorporated patent documents.

Isolated Primer Pairs

[0414] In some embodiments, the invention provides isolated nucleic acids that, when used as primers in a polymerase chain reaction, amplify a subject polynucleotide, or a polynucleotide containing a subject polynucleotide. The amplified polynucleotide is from about 20 to about 50, from about 50 to about 75, from about 75 to about 100, from about 100 to about 125, from about 125 to about 150, from about 150 to about 175, from about 175 to about 200, from about 200 to about 250, from about 250 to about 300, from about 300 to about 350, from about 350 to about 400, from about 400 to about 500, from about 500 to about 600, from about 600 to about 700, from about 700 to about 800, from about 800 to about 900, from about 900 to about 1000, from about 1000 to about 2000, from about 2000 to about 3000, from about 3000 to about 4000, from about 4000 to about 5000, or from about 5000 to about 6000 nucleotides or more in length.

[0415] The isolated nucleic acids themselves are from about 10 to about 20, from about 20 to about 30, from about 30 to about 40, from about 40 to about 50, from about 50 to about 100, or from about 100 to about 200 nucleotides in length. Generally, the nucleic acids are used in pairs in a polymerase chain reaction, where they are referred to as "forward" and "reverse" primers.

[0416] Thus, in some embodiments, the invention provides a pair of isolated nucleic acid molecules, each from about 10 to about 200 nucleotides in length, the first nucleic acid molecule of the pair comprising a sequence of at least 10 contiguous nucleotides having 100% sequence identity to a nucleic acid sequence as shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697 and the second nucleic acid molecule of the pair comprising a sequence of at least 10 contiguous nucleotides having 100% sequence identity to the reverse complement of the nucleic acid sequence shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697 , wherein the sequence of the second nucleic acid molecule is located 3' of the nucleic acid sequence of the first nucleic acid molecule shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697. The primer nucleic

acids are prepared using any known method, e.g., automated synthesis, and can be chosen to specifically amplify a cDNA copy of an mRNA encoding a subject polypeptide.

[0417]   In some embodiments, the first and/or the second nucleic acid molecules comprise a detectable label. The label can be a radioactive molecule, fluorescent molecule or another molecule, e.g., hapten, as described in detail above. Further, the label can be a two stage system, where the amplified DNA is conjugated to another molecule, i.e., biotin, digoxin, or a hapten, that has a high affinity binding partner, i.e., avidin, antidigoxin, or a specific antibody, respectively, and the binding partner conjugated to a detectable label. The label can be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

[0418]   Conditions that increase stringency of both DNA/DNA and DNA/RNA hybridization reactions are widely known and published in the art. See, for example, Sambrook, 1989, and examples provided above. Examples of relevant conditions include (in order of increasing stringency): incubation temperatures of 25°C, 37°C, 50°C, and 68°C; buffer concentrations of 10 x SSC, 6 x SSC, 1 x SSC, 0.1 x SSC (where 1 x SSC is 0.15 M NaCl and 15 mM citrate buffer); and their equivalents using other buffer systems; formamide concentrations of 0%, 25%, 50%, and 75%; incubation times from 5 minutes to 24 hours; 1, 2, or more washing steps; wash incubation times of 1, 2, or 15 minutes; and wash solutions of 6 x SSC, 1 x SSC, 0.1 x SSC, or deionized water.

[0419]   For example, "high stringency conditions" include hybridization in 50% formamide, 5X SSC, 0.2 µg/µl poly(dA), 0.2 µg/µl human cot1 DNA, and 0.5% SDS, in a humid oven at 42°C overnight, followed by successive washes in 1X SSC, 0.2% SDS at 55°C for 5 minutes, followed by washing at 0.1X SSC, 0.2% SDS at 55°C for 20 minutes. Further examples of high stringency conditions include hybridization at 50°C and 0.1×SSC (15 mM sodium chloride/1.5 mM sodium citrate); overnight incubation at 42°C in a solution containing 50% formamide, 1 × SSC (150 mM NaCl, 15 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5 × Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1 × SSC at about 65°. High stringency conditions also include aqueous hybridization (e.g., free of formamide) in

6X SSC (where 20X SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% sodium dodecyl sulfate (SDS) at 65°C for about 8 hours (or more), followed by one or more washes in 0.2 X SSC, 0.1% SDS at 65°C. Highly stringent hybridization conditions are hybridization conditions that are at least as stringent as any one of the above representative conditions. Other stringent hybridization conditions are known in the art and can also be employed to identify nucleic acids of this particular embodiment of the invention.

[0420]   Conditions of "reduced stringency," suitable for hybridization to molecules encoding structurally and functionally related proteins, or otherwise serving related or associated functions, are the same as those for high stringency conditions but with a reduction in temperature for hybridization and washing to lower temperatures (e.g., room temperature or about 22°C to 25°C). For example, moderate stringency conditions include aqueous hybridization (e.g., free of formamide) in 6X SSC, 1% SDS at 65°C for about 8 hours (or more), followed by one or more washes in 2X SSC, 0.1% SDS at room temperature. Low stringency conditions include, for example, aqueous hybridization at 50°C and 6×SSC (0.9 M sodium chloride/0.09 M sodium citrate) and washing at 25°C in 1×SSC (0.15 M sodium chloride/0.015 M sodium citrate).

[0421]   The specificity of a hybridization reaction allows any single-stranded sequence of nucleotides to be labeled with a radioisotope or chemical and used as a probe to find a complementary strand, even in a cell or cell extract that contains millions of different DNA and RNA sequences. Probes of this type are widely used to detect the nucleic acids corresponding to specific genes, both to facilitate the purification and characterization of the genes after cell lysis and to localize them in cells, tissues, and organisms.

[0422]   Moreover, by carrying out hybridization reactions under conditions of "reduced stringency," a probe prepared from one gene can be used to find homologous evolutionary relatives - both in the same organism, where the relatives form part of a gene family, and in other organisms, where the evolutionary history of the nucleotide sequence can be traced. A person skilled in the art would recognize how to modify the conditions to achieve the requisite degree of stringency for a particular hybridization.

Libraries

[0423] The polynucleotide libraries of the invention generally comprise a collection of sequence information of a plurality of polynucleotide sequences, where at least one of the polynucleotides has a sequence shown in SEQ ID NOS.: 1 -1231 and 2463 - 3697. By plurality is meant at least 2, at least 3, or at least all of the sequences in the Sequence Listing. The information may be provided in either biochemical form (e.g., as a collection of polynucleotide molecules), or in electronic form (e.g., as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer-based system, a computer data file, and/or as a part of a computer program). The length and number of polynucleotides in the library will vary with the nature of the library, e.g., if the library is an oligonucleotide array, a cDNA array, or a computer database of the sequence information.

[0424] The sequence information contained in either a biochemical or an electronic library of polynucleotides can be used in a variety of ways, e.g., as a resource for gene discovery, as a representation of sequences expressed in a selected cell type (e.g., cell type markers), or as markers of a given disorder or disease state. In general, a disease marker is a representation of a gene product that is present in all cells affected by disease either at an increased or decreased level relative to a normal cell (e.g., a cell of the same or similar type that is not substantially affected by disease). For example, a polynucleotide sequence in a library can be a polynucleotide that represents an mRNA, polypeptide, or other gene product encoded by the polynucleotide, that is either over-expressed or under-expressed in one cell compared to another (e.g., a first cell type compared to a second cell type; a normal cell compared to a diseased cell; a cell not exposed to a signal or stimulus compared to a cell exposed to that signal or stimulus; and the like).

[0425] The nucleotide sequence information of the library can be embodied in any suitable form, e.g., electronic or biochemical forms. For example, a library of sequence information embodied in electronic form comprises an accessible computer data file that may contain the representative nucleotide sequences of genes that are differentially expressed (e.g., over-expressed or under-expressed) as between, e.g., a first cell type compared to a second cell type (e.g., expression in a brain cell compared to expression in a kidney cell); a normal cell compared to a diseased cell (e.g., a non-cancerous cell compared to a cancerous cell); a cell not exposed to an internal or external signal or stimulus compared to a cell exposed to that signal or stimulus (e.g.,

a cell contacted with a ligand compared to a control cell not contacted with the ligand); and the like. Other combinations and comparisons of cells will be readily apparent to the ordinarily skilled artisan. Biochemical embodiments of the library include a collection of nucleic acid molecules that have the sequences of the genes in the library, where the nucleic acids can correspond to the entire gene in the library or to a fragment thereof, as described in greater detail below.

[0426] Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. For example, the nucleic acid sequences of any of the polynucleotides shown in SEQ ID NOS.: 1 -1231 and 2463 - 3697 can be recorded on computer readable media of a computer-based system, e.g., any medium that can be read and accessed directly by a computer. One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, e.g., word processing text file, database format, etc. In addition to the sequence information, electronic versions of the libraries of the invention can be provided in conjunction or connection with other computer-readable information and/or other types of computer-based files (e.g., searchable files, executable files, etc, including, but not limited to, for example, search program software, etc.).

[0427] By providing the nucleotide sequence in computer readable form in a computer-based system, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. Conventional bioinformatics tools can be utilized to analyze sequences to determine sequence identity, sequence similarity, and gap information. For example, the gapped BLAST (Altschul et al., 1990, Altschul et al., 1997), and BLAZE (Brutlag et al., 1993) search algorithms on a Sybase system, or the TeraBLAST (TimeLogic, Crystal Bay, Nevada) program optionally running on a specialized computer platform available from TimeLogic, can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms. Homology between sequences of interest can be determined using the local homology algorithm of Smith and Waterman, 1981, as well as the BestFit program (Rechid et al., 1989), and the FastDB algorithm (FastDB, 1988; described in Current Methods in Sequence

Comparison and Analysis, Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp. 127-149, 1988, Alan R. Liss, Inc).

[0428] Alignment programs that permit gaps in the sequence include Clustalw (Thompson et al., 1994), FASTA3 (Pearson, 2000) Align0 (Myers and Miller, 1988), and TCoffee (Notredame et al., 2000). Other methods for comparing and aligning nucleotide and protein sequences include, for example, BLASTX (NCBI), the Wise package (Birney and Durbin, 2000), and FASTX (Pearson, 2000). These algorithms determine sequence homology between nucleotide and protein sequences without translating the nucleotide sequences into protein sequences. Other techniques for alignment are also known in the art (Doolittle, et al., 1996; BLAST, available from the National Center for Biotechnology Information; FASTA, available in the Genetics Computing Group (GCG) package, from Madison, Wisconsin, USA, a wholly owned subsidiary of Oxford Molecular Group, Inc.; Schlessinger, 1988a; Schlessinger, 1988b; and Needleman and Wunch, 1970).

[0429] Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, etc. The reference sequence is usually at least about 18 nt long, at least about 30 nt long, or may extend to the complete sequence that is being compared.

[0430] One parameter for determining percent sequence identity is the percentage of the alignment in the region of strongest alignment between a target and a query sequence. Methods for determining this percentage involve, for example, counting the number of aligned bases of a query sequence in the region of strongest alignment and dividing this number by the total number of bases in the region. For example, 10 matches divided by 11 total residues gives a percent sequence identity of approximately 90.9%. The length of the aligned region is typically at least about 55%, at least about 58%, or at least about 60% of the total sequence length, and can be as great as about 62%, as great as about 64%, and even as great as about 66% of the total sequence length.

[0431] The present invention includes human and mouse polynucleotide and polypeptide sequences that are at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% homologous to the sequences in the Sequence Listing, based on using the method of determining sequence identity with the insertion of gaps to detect the maximum degree of sequence identity. In other

embodiments of interest, homology will be at least about 80%, at least about 85%, or as high as about 90%.

[0432]   A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks the relative expression levels of different polynucleotides. Such presentation provides a skilled artisan with a ranking of relative expression levels to determine a gene expression profile.

[0433]   As discussed above, the library of the invention also encompasses biochemical libraries of the polynucleotides shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697, e.g., collections of nucleic acids representing the provided polynucleotides. The biochemical libraries can take a variety of forms, e.g., a solution of cDNAs, a pattern of probe nucleic acids stably associated with a surface of a solid support (i.e., an array) and the like. Of particular interest are nucleic acid arrays in which one or more of the polynucleotide sequences shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697 is represented on the array. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis, and the like, as disclosed in the herein-listed exemplary patent documents.

[0434]   In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the polypeptides of the library will represent at least a portion of the polypeptides encoded by a gene corresponding to one or more of the sequences shown in SEQ ID NOS.: 1 -1231 and 2463 - 3697.

[0435]   Further, analogous libraries of antibodies are also provided, where the libraries comprise antibodies or fragments thereof that specifically bind to at least a portion of at least one of the subject polypeptides. Further, antibody libraries may comprise antibodies or fragments thereof that specifically inhibit binding of a subject polypeptide to its ligand or substrate, or that specifically inhibit binding of a subject polypeptide as a substrate to another molecule. Moreover, corresponding nucleic acid libraries are also provided, comprising polynucleotide sequences that encode the antibodies or antibody fragments described above.

Polypeptides

Sequences

[0436] This invention provides novel polypeptides, and related polypeptide compositions. The novel polypeptides of the invention encompass proteins with amino acid sequences as shown in SEQ ID NOS.: 1232 - 2462, or encoded by the nucleic acids having nucleotide sequences shown in SEQ ID NOS.: 1 -1231 and 2463 - 3697. The subject polypeptides are human polypeptides, fragments thereof, variants (such as splice variants), homologs from other species, and derivatives thereof. In particular embodiments, a polypeptide of the invention has an amino acid sequence substantially identical to the sequence of any polypeptide encoded by a polynucleotide sequence shown in SEQ ID NOS.: 1 -1231 and 2463 - 3697.

[0437] These polypeptides may reside within the cell, or extracellularly. They may be secreted from the cell, reside in the cytoplasm, in the membranes, or in any of the intracellular organelles, including the nucleus, mitochondria, ribosomes, or storage granules.

[0438] In many embodiments, a novel polypeptide of the invention functions as a secreted protein, a single-transmembrane protein, a multiple-transmembrane protein, a kinase, a protein kinase, a ligase, a nuclear hormone receptor, a phosphatase, a protease, a phosphodiesterase, a kinesin, an immunoglobulin, a T-cell receptor, or a glycosylphosphatidylinositol anchor. A novel polypeptide of the invention can also possess one or more of the following functions or properties: (1) an activator functioning to regulate one or more genes by increasing the rate of transcription, (2) an activator functioning to positively modulate an allosteric enzyme, (3) an adaptor functioning to sort cargo molecules into transport vesicles, (4) an adaptor functioning to form a clathrin-coated vesicle, (5) an adhesion molecule functioning to mediate the adhesion of cells with other cells and/or the extracellular matrix, (6) an ATPase functioning to move ions or small molecules across a membrane against a chemical concentration gradient or electrical potential, (7) an ATPase functioning to translocate nucleotides across membranes, (8) a breakpoint-related sequence functioning as an oncoprotein, (9) a breakpoint-related sequence functioning as a tumor-specific antigen, (10) a channel functioning as a water channel, (11) a channel functioning as an ion channel, (12) a checkpoint-related sequence functioning at DNA damage checkpoints, (13) a checkpoint-related sequence functioning at replication checkpoints, (14) a checkpoint-related sequence functioning to initiate signal transduction cascades eliciting cell cycle arrest, DNA repair, or apoptosis, (15) a complex functioning as a protein scaffold, (16) a complex

functioning in ADP-ribosylation, (17) a dehydrogenase functioning to synthesize amino acids, (18) a disintegrin functioning to inhibit blood clotting, (19) a disintegrin functioning as a metallopeptidase, (20) a GTPase functioning as a negative regulator of p53, (21) a GTPase functioning to stimulate ras GTPase activity, (22) a helicase functioning in DNA replication, (23) a hydrolase functioning in proprionate metabolism, (24) an integrase functioning to integrate a DNA copy of a retroviral genome into a host chromosome, (25) an integrin functioning as a tumor marker, (26) an integrin functioning in cell migration, (27) an isomerase functioning as an immunosuppressant, (28) a membrane protein functioning as a scaffolding component at the cytoplasmic face of a lipid raft, (29) a membrane protein functioning as a ligand for a receptor tyrosine kinase, (30) oxygenases and peroxidases functioning as antioxidants, (31) a phospholipase functioning in eicosanoid synthesis, (32) a phospholipase functioning in preserving the intestinal mucosa, (33) a prosaposin functioning in lipid catabolism, (34) a proteasome component functioning in muscle wasting, (35) a reductase-related sequence functioning as a coenzyme A reductase inhibitor, (36) a reverse transcriptase functioning as an RNA-dependent reverse transcriptase, (37) a reverse transcriptase functioning as a DNA-dependent reverse transcriptase, (38) an RNase functioning in viral assembly, (39) an RNase H functioning to form oligonucleotides that prime DNA synthesis, (40) an RNase H functioning to cleave the RNA strand of an RNA-DNA hybrid, (41) SH3 domains functioning in actin cytoskeletal organization, (42) SH3 domains functioning in signal transduction, (43) a synthetase functioning as an autoantigen (44) synthetases functioning in nucleotide sugar phosphate synthesis, (45) TATA boxes functioning as a transcription initiators, (46) tat functioning as a transcriptional coactivator, (47) transferases functioning in signal transduction, (48) transposases functioning as gene transfer agents, (49) ubiquitins functioning to protect cells against tumor necrosis factor induced cell death, (50) proteasome components and ubiquitin functioning in protein degradation, (51) a virus-related sequence functioning to confer resistance to infection by viruses, (52) other sequences of the invention interacting with one or more proteins, (53) other sequences of the invention enzymatically modifying one or more proteins, (54) other sequences of the invention binding one or more small molecule ligands, (55) other sequences of the invention binding one or more peptides, (56) other sequences of the invention binding one or more carbohydrates, and (57) other sequences of the invention functioning in vesicular transport.

[0439]   In some embodiments, the present novel polypeptide modulates the cells or tissues of animals, particularly humans, such as, for example, by stimulating, enhancing or inhibiting T or B cell function or the function of other hematopoeitic cells or bone marrow cells; modulates adult or embryonic stem cell or precursor cell growth or differentiation; modulates cell function or activity of neuronal cells or other cells of the CNS, heart cells, liver cells, kidney cells, lung cells, pancreatic cells, gastrointestinal cells, spleen cells, breast cells, prostate cells, ovarian cells, and the like.

[0440]   In some embodiments, a subject polypeptide is present as a multimer. Multimers include homodimers, homotrimers, homotetramers, and multimers that include more than four monomeric units. Multimers also include heteromultimers, e.g., heterodimers, heterotrimers, heterotetramers, etc. where the subject polypeptide is present in a complex with proteins other than the subject polypeptide. Where the multimer is a heteromultimer, the subject polypeptide can be present in a 1:1 ratio, a 1:2 ratio, a 2:1 ratio, or other ratio, with the other protein(s).

[0441]   In addition to the above specifically listed proteins, polypeptides from other species are also provided, including mammals, such as: primates, rodents, e.g., mice, rats, hamsters, guinea pigs; domestic animals, e.g., sheep, pig, horse, cow, goat, rabbit, dog, cat; and humans, as well as non-mammalian species, e.g., avian, reptile and amphibian, insect, crustacean, fish, plant, fungus, and protozoa.

[0442]   By "homolog" is meant a protein having at least about 35 %, at least about 40%, at least about 60%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, or at least about 95%, or higher, amino acid sequence identity to the reference polypeptide, as measured with the "GAP" program (part of the Wisconsin Sequence Analysis Package available through the Genetics Computer Group, Inc. (Madison WI)), where the parameters are: Gap weight:12; length weight:4. In many embodiments of interest, homology will be at least about 75%, at least about 80%, or at least 85%, where in certain embodiments of interest, homology will be as high as about 90%.

[0443]   Also provided are polypeptides that are substantially identical to the at least one amino acid sequence shown in the Sequence Listing, or a fragment thereof, whereby substantially identical is meant that the protein has an amino acid sequence identity to the reference sequence of at least about 75%, at least about 80%,

at least about 85%, at least about 90%, at least about 95%, at least about 97%, at least about 98%, or at least about 99%.

[0444]   The proteins of the subject invention (e.g., polypeptides encoded by the nucleotide sequences shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697, and polypeptide sequences shown in SEQ ID NOS.: 1232 - 2462) have been separated from their naturally occurring environment and are present in a non-naturally occurring environment. In certain embodiments, the proteins are present in a composition where they are more concentrated than in their naturally occurring environment. For example, purified polypeptides are provided.

[0445]   In addition to naturally occurring proteins, polypeptides that vary from naturally occurring forms are also provided. Fusion proteins can comprise a subject polypeptide, or fragment thereof, and a polypeptide other than a subject polypeptide ("the fusion partner") fused in-frame at the N-terminus and/or C-terminus of the subject polypeptide, or internally to the subject polypeptide.

[0446]   Suitable fusion partners include, but are not limited to, immunologically detectable proteins (e.g., epitope tags, such as hemagglutinin, FLAG, and c-myc); polypeptides that provide a detectable signal or that serve as detectable markers (e.g., a fluorescent protein, e.g., a green fluorescent protein, a fluorescent protein from an Anthozoan species; β-galactosidase; luciferase; cre recombinase; and the like); polypeptides that provide a catalytic function or induce a cellular response; polypeptides that provide for secretion of the fusion protein from a eukaryotic cell; polypeptides that provide for secretion of the fusion protein from a prokaryotic cell; polypeptides that provide for binding to metal ions (e.g., His_n, where n = 3-10, e.g., 6His) and structural proteins. Fusion partners can also be those that are able to stabilize the present polypeptide, such as polyethylene glycol ("PEG") and a fragment of an immunoglobulin, such as the Fc fragment of IgG, IgE, IgA, IgM, and/or IgD.

[0447]   Detection methods are chosen based on the detectable fusion partner. For example, where the fusion partner provides an immunologically recognizable epitope, an epitope-specific antibody can be used to quantitatively detect the level of polypeptide. In some embodiments, the fusion partner provides a detectable signal, and in these embodiments, the detection method is chosen based on the type of signal generated by the fusion partner. For example, where the fusion partner is a fluorescent protein, fluorescence is measured.

[0448]   Where the fusion partner is an enzyme that yields a detectable product, the product can be detected using an appropriate means. For example, β-galactosidase can, depending on the substrate, yield a colored product that can be detected with a spectrophotometer, and the fluorescent protein luciferase can yield a luminescent product detectable with a luminometer.

[0449]   In some embodiments, a polypeptide of the invention comprises at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, at least about 900, at least about 950, or at least about 1000 contiguous amino acid residues of at least one of the sequences according to SEQ ID NOS.: 1232 - 2462, up to and including the entire amino acid sequence.

[0450]   Fragments of the subject polypeptides, as well as polypeptides comprising such fragments, are also provided. Fragments of polypeptides of interest will typically be at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, or at least 300 aa in length or longer, where the fragment will have a stretch of amino acids that is identical to the subject protein of at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, or at least about 50 aa in length.

[0451]   In some embodiments, fragments exhibit one or more activities associated with a corresponding naturally occurring polypeptide. Fragments find utility in generating antibodies to the full-length polypeptide; and in methods of screening for candidate agents that bind to and/or modulate polypeptide activity. Specific fragments of interest include those with enzymatic activity, those with biological activity including the ability to serve as an epitope or immunogen, and fragments that bind to other proteins or to nucleic acids.

[0452]   The invention provides polypeptides comprising such fragments, including, e.g., fusion polypeptides comprising a subject polypeptide fragment fused in frame (directly or indirectly) to another protein (the "fusion partner"), such as the

signal peptide of one protein being fused to the mature polypeptide of another protein. Such fusion proteins are typically made by linking the encoding polynucleotides together in a vector or cassette. Suitable fusion partners include, but are not limited to, immunologically detectable proteins (e.g., epitope tags, such as hemagglutinin, FLAG, and c-myc); polypeptides that provide a detectable signal or that serve as detectable markers (e.g., a fluorescent protein, e.g., a green fluorescent protein, a fluorescent protein from an Anthozoan species; β-galactosidase; luciferase; cre recombinase); polypeptides that provide a catalytic function or induce a cellular response; polypeptides that provide for secretion of the fusion protein from a eukaryotic cell; polypeptides that provide for secretion of the fusion protein from a prokaryotic cell; polypeptides that provide for binding to metal ions (e.g., His_n, where $n = 3-10$, e.g., 6His) and structural proteins. Fusion partners can also be those that are able to stabilize the present polypeptide, such as polyethylene glycol ("PEG") and a fragment of an immunoglobulin, such as the Fc fragment of IgG, IgE, IgA, IgM, and/or IgD.

Polypeptide Preparation.

[0453] Polypeptides of the invention can be obtained from naturally-occurring sources or produced synthetically. The sources of naturally occurring polypeptides will generally depend on the species from which the protein is to be derived, i.e., the proteins will be derived from biological sources that express the proteins. The subject proteins can also be derived from synthetic means, e.g., by expressing a recombinant gene encoding a protein of interest in a suitable system or host or enhancing endogenous expression, as described in more detail above. Further, small peptides can be synthesized in the laboratory by techniques well known in the art.

[0454] In all cases, the product can be recovered by any appropriate means known in the art. For example, convenient protein purification procedures can be employed (e.g., see <u>Guide to Protein Purification</u>, Deuthscher et al., 1990). That is, a lysate can be prepared from the original source, (e.g., a cell expressing endogenous polypeptide, or a cell comprising the expression vector expressing the polypeptide(s)), and purified using HPLC, exclusion chromatography, gel electrophoresis, or affinity chromatography, and the like.

[0455] The invention thus also provides methods of producing polypeptides. Briefly, the methods generally involve introducing a nucleic acid construct into a host

cell *in vitro* and culturing the host cell under conditions suitable for expression, then harvesting the polypeptide, either from the culture medium or from the host cell, (e.g., by disrupting the host cell), or both, as described in detail above. The invention also provides methods of producing a polypeptide using cell-free *in vitro* transcription/translation methods, which are well known in the art, also as provided above

[0456]   Moreover, the invention provides polypeptides, including polypeptide fragments, as targets for therapeutic intervention, including use in screening assays, for identifying agents that modulate polypeptide level and/or activity, and as targets for antibody and small molecule therapeutics, for example, in the treatment of disorders.

Methods

[0457]   The present invention provides methods of producing a subject polypeptide and provides antibodies that specifically bind to a subject polypeptide. The present invention further provides screening methods for identifying agents that modulate a level or an activity of a subject polypeptide or polynucleotide. The present invention thus also provides agents that modulate a level or an activity of a subject polypeptide or polynucleotide, as well as compositions, including pharmaceutical compositions, comprising a subject agent.

[0458]   The present invention further provides methods for treating disorders such as, for example, cancer and other proliferative disorders, inflammatory and immune disorders, metabolic disorders, and bacterial or viral disorders.

Diagnostic and Therapeutic Applications

Screening and Diagnostic Methods

1. Identifying Biological Molecules that Interact with a Polypeptide

[0459]   Formation of a binding complex between a subject polypeptide and an interacting polypeptide or other macromolecule (e.g., DNA, RNA, lipids, polysaccharides, and the like) can be detected using any known method. Suitable methods include: a yeast two-hybrid system (Zhu et al., 1997; Fields and Song, 1989; U.S. Pat. No. 5,283,173; Chien et al. 1991); a mammalian cell two-hybrid method; a fluorescence resonance energy transfer (FRET) assay; a bioluminescence resonance energy transfer (BRET) assay; a fluorescence quenching assay; a fluorescence anisotropy assay (Jameson and Sawyer, 1995); an immunological assay; and an assay involving binding of a detectably labeled protein to an immobilized protein.

[0460] Immunological assays, and assays involving binding of a detectably labeled protein to an immobilized protein can be performed in a variety of ways. For example, immunoprecipitation assays can be designed such that the complex of protein and an interacting polypeptide is detected by precipitation with an antibody specific for either the protein or the interacting polypeptide.

[0461] FRET detects formation of a binding complex between a subject polypeptide and an interacting polypeptide. It involves the transfer of energy from a donor fluorophore in an excited state to a nearby acceptor fluorophore. For this transfer to take place, the donor and acceptor molecules must be in close proximity (e.g., less than 10 nanometers apart, usually between 10 and 100 Å apart), and the emission spectra of the donor fluorophore must overlap the excitation spectra of the acceptor fluorophore. In these embodiments, a fluorescently labeled subject protein serves as a donor and/or acceptor in combination with a second fluorescent protein or dye.

[0462] Fluorescent proteins can be produced by generating a construct comprising a protein and a fluorescent fusion partner. These are well-known in the art, as described above, including green fluorescent protein (GFP), i.e., a "humanized" version of a GFP, e.g., wherein codons of the naturally-occurring nucleotide sequence are changed to more closely match human codon bias; a GFP derived from *Aequoria victoria* or a derivative thereof, e.g., a "humanized" derivative such as Enhanced GFP, which are available commercially, e.g., from Clontech, Inc.; other fluorescent mutants of a GFP from *Aequoria victoria*, e.g., as described in U.S. Patent No. 6,066,476; 6,020,192; 5,985,577; 5,976,796; 5,968,750; 5,968,738; 5,958,713; 5,919,445; 5,874,304; a GFP from another species such as *Renilla reniformis, Renilla mulleri*, or *Ptilosarcus guernyi*, as previously described (WO 99/49019; Peelle et al., 2001), "humanized" recombinant GFP (hrGFP) (Stratagene®); any of a variety of fluorescent and colored proteins from Anthozoan species, (e.g., Matz et al., 1999); as well as proteins labeled with other fluorescent dyes, fluorescein and it derivatives, e.g., fluorescein isothiocyanate (FITC), 6-carboxyfluorescein (6-FAM), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE); rhodamine dyes, e.g., Texas red, phycoerythrin, tetramethylrhodamine, rhodamine, 6-carboxy-X-rhodamine (ROX); coumarin and its derivatives, e.g., 7-amino-4-methylcoumarin, aminocoumarin; bodipy dyes, such as Bodipy FL; cascade blue; Oregon green; eosins

and erythrosins; cyanine dyes, e.g., allophycocyanin, Cy3, Cy5, and N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); macrocyclic chelates of lanthanide ions, e.g., quantum dye, etc; and chemiluminescent molecules, e.g., luciferases.

[0463] Fluorescent subject proteins can also be generated by producing the subject protein in an auxotrophic strain of bacteria which requires addition of one or more amino acids in the medium for growth. A subject protein-encoding construct that provides for expression in bacterial cells is introduced into the auxotrophic strain, and the bacteria are cultured in the presence of a fluorescent amino acid, which is incorporated into the subject protein produced by the bacterium. The subject protein is then purified from the bacterial culture using standard methods for protein purification.

[0464] BRET is a protein-protein interaction assay based on energy transfer from a bioluminescent donor to a fluorescent acceptor protein. The BRET signal is measured by the ratio of the amount of light emitted by the acceptor to the amount of light emitted by the donor. The ratio of these two values increases as the two proteins are brought into proximity. The BRET assay has been described in the literature (U.S. Patent Nos. 6,020,192; 5,968,750; 5,874,304; Xu, et al. 1999). BRET assays can be performed by analyzing transfer between a bioluminescent donor protein and a fluorescent acceptor protein. Interaction between the donor and acceptor proteins can be monitored by a change in the ratio of light emitted by the bioluminescent and fluorescent proteins. In this application, the subject protein serves as donor and/or acceptor protein.

[0465] Fluorescence anisotropy is a measurement of the rotational mobility of a multi-molecular complex. It can be used to generate information about the binding of one molecule to another, including the affinity and specificity of binding sites. It can be applied to polypeptides or nucleic acids of the present invention.

[0466] Fluorescence quenching measurements are useful in detecting protein multimerization, such as where the subject protein interacts with at least a second protein and, for example, where multimerization interaction is affected by a test agent. As used herein, the term "multimerization" refers to formation of dimers, trimers, tetramers, and higher multimers of the subject protein. Whether a subject protein forms a complex with one or more additional protein molecules can be determined using any known assay, including assays as described above for interacting proteins. Formation of multimers can also be detected using non-denaturing gel

This page is not part of

the document!

US2003026780 / 2004-035732

2/4

Date: Apr 29, 2004

Recipient: IB
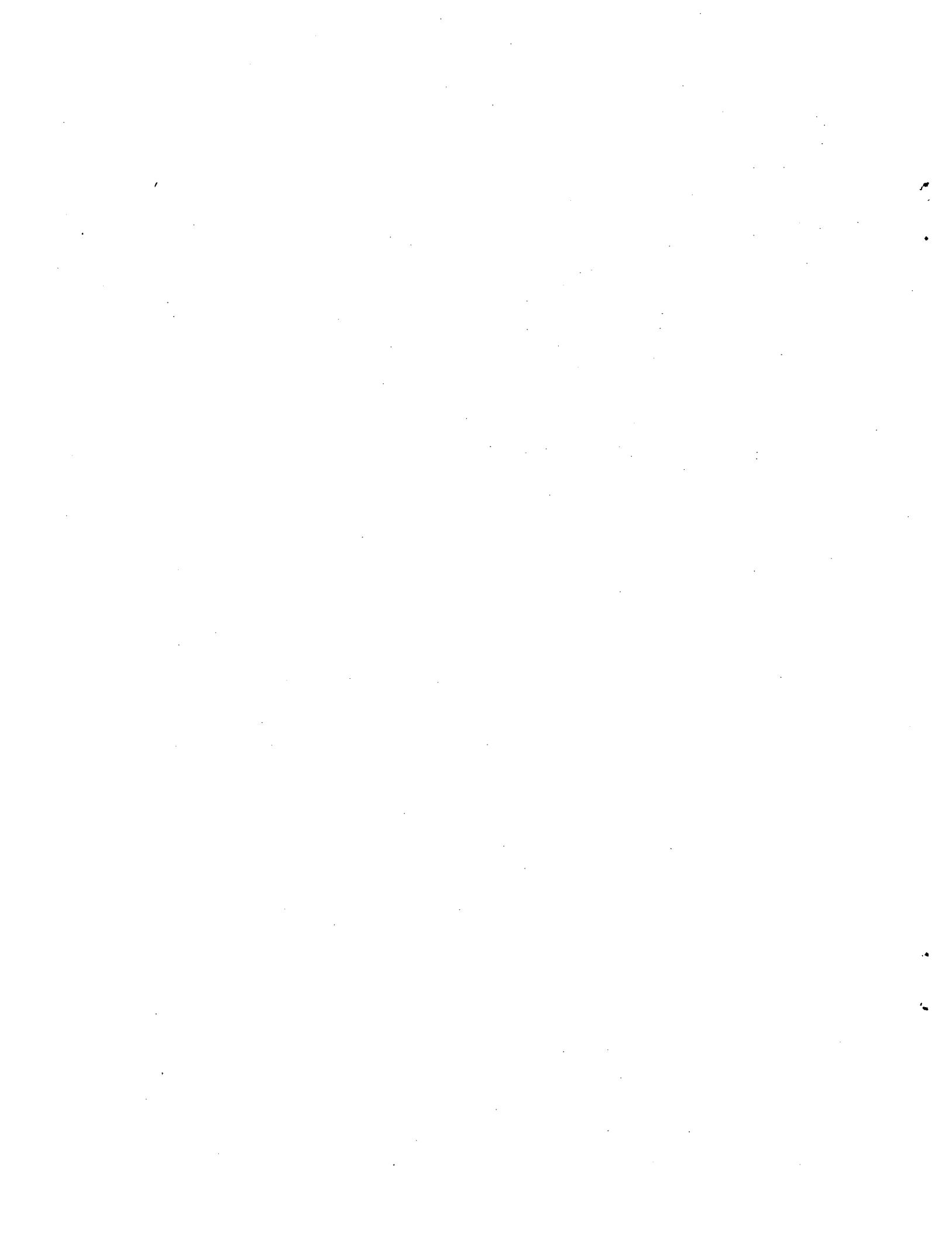
electrophoresis, where multimerized subject protein migrates more slowly than monomeric subject protein. Formation of multimers can also be detected using fluorescence quenching techniques.

[0467] Formation of multimers can also be detected by analytical ultracentrifugation, for example through glycerol or sucrose gradients, and subsequent visualization of a subject protein in gradient fractions by Western blotting or staining of SDS-polyacrylamide gels. Multimers are expected to sediment at defined positions in such gradients. Formation of multimers can also be detected using analytical gel filtration, e.g., in HPLC or FPLC systems, e.g., on columns such as Superdex 200 (Pharmacia Amersham Inc.). Multimers run at defined positions on these columns, and fractions can be analyzed as above. The columns are highly reproducible, allowing one to relate the number and position of peaks directly to the multimerization status of the protein.

2. Detecting mRNA Levels and Monitoring Gene Expression

[0468] The present invention provides methods for detecting the presence of mRNA in a biological sample. The methods can be used, for example, to assess whether a test compound affects gene expression, either directly or indirectly. The present invention provides diagnostic methods to compare the abundance of a nucleic acid with that of a control value, either qualitatively or quantitatively, and to relate the value to a normal or abnormal expression pattern.

[0469] Methods of measuring mRNA levels are known in the art (Pietu, 1996; Zhao, 1995; Soares, 1997; Raval, 1994; Chalifour, 1994; Stolz, 1996; Hong, 1982; McGraw, 1984; WO 97/27317). These methods generally comprise contacting a sample with a polynucleotide of the invention under conditions that allow hybridization and detecting hybridization, if any, as an indication of the presence of the polynucleotide of interest. Appropriate controls include the use of a sample lacking the polynucleotide mRNA of interest, or the use of a labeled polynucleotide of the same "sense" as a polynucleotide mRNA of interest. Detection can be accomplished by any known method, including, but not limited to, *in situ* hybridization, PCR, RT-PCR, and "Northern" or RNA blotting, or combinations of such techniques, using a suitably labeled subject polynucleotide. A variety of labels and labeling methods for polynucleotides are known in the art and can be used in the assay methods of the invention. A common method employed is use of microarrays

135

which can be purchased or customized, for example, through conventional vendors such as Affymetrix.

[0470]   In some embodiments, the methods involve generating a cDNA copy of an mRNA molecule in a biological sample, and amplifying the cDNA using an isolated primer pairs as described above, i.e., a set of two nucleic acid molecules that serve as forward and reverse primers in an amplification reaction (e.g., a polymerase chain reaction). The primer pairs are chosen to specifically amplify a cDNA copy of an mRNA encoding a polypeptide. A detectable label can be included in the amplification reaction, as provided above. Methods using PCR amplification can be performed on the DNA from a single cell, although it is convenient to use at least about 10^5 cells.

[0471]   The present invention provides methods for monitoring gene expression. Changes in a promoter or enhancer sequence that can affect gene expression can be examined in light of expression levels of the normal allele by various methods known in the art. Methods for determining promoter or enhancer strength include quantifying the expressed natural protein, and inserting the variant control element into a vector with a quantitative reporter gene such as β-galactosidase, luciferase, or chloramphenicol acetyltransferase (CAT).

3. Detecting Polymorphisms and Mutations

[0472]   Biochemical studies can determine whether a sequence polymorphism in a coding region or control region is associated with disease. Disease-associated polymorphisms can include deletion or truncation of the gene, mutations that alter expression level, or mutations that affect protein function, etc. A number of methods are available to analyze nucleic acids for the presence of a specific sequence, e.g., a disease associated polymorphism. Genomic DNA can be used when large amounts of DNA are available. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express the gene provide a source of mRNA, which can be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid can be amplified by conventional techniques, i.e., PCR, to provide sufficient amounts for analysis. (Saiki et al., 1988; Sambrook et al., 1989, pp.14.2-14.33). Alternatively, various methods are known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms (Riley et al., 1990; Delahunty et al., 1996).

[0473]　The sample nucleic acid, e.g., an amplified or cloned fragment, is analyzed by one of a number of methods known in the art. The nucleic acid can be sequenced by dideoxy nucleotide sequencing, or other methods, and the sequence of bases compared to a wild-type sequence. Hybridization with the variant sequence can also be used to determine its presence, e.g., by Southern blots, dot blots, etc. The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilized on a solid support, as described in US Pat. No. 5,445,934, or WO 95/35505, can also be used as a means of detecting the presence of variant sequences. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices can detect variation as alterations in electrophoretic mobility resulting from conformational changes created by DNA sequence alterations. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease, the sample can be digested with that endonuclease, and the products fractionated according to their size to determine whether the fragment was digested. Fractionation can be performed by gel or capillary electrophoresis, for example with acrylamide or agarose gels.

[0474]　Screening for mutations in a gene can be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting deletions that might affect the biological activity of the protein. Various immunoassays designed to detect polymorphisms in proteins can be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to be effective screening tools. The activity of the encoded protein can be determined by comparison with the wild-type protein.

4. Detecting and Monitoring Polypeptide Presence and Biological Activity

[0475]　The present invention provides methods for detecting the presence and/or biological activity of a subject polypeptide in a biological sample. The assay used will be appropriate to the biological activity of the particular polypeptide. Thus, e.g., where the biological activity is an enzymatic activity, the method will involve contacting the sample with an appropriate substrate, and detecting the product of the enzymatic reaction on the substrate. Where the biological activity is binding to a second macromolecule, the assay detects protein-protein binding, protein-DNA binding, protein-carbohydrate binding, or protein-lipid binding, as appropriate, using well known assays. Where the biological activity is signal transduction (e.g.,

137

transmission of a signal from outside the cell to inside the cell) or transport, an appropriate assay is used, such as measurement of intracellular calcium ion concentration, measurement of membrane conductance changes, or measurement of intracellular potassium ion concentration.

[0476] The present invention also provides methods for detecting the presence or measuring the level of a normal or abnormal polypeptide in a biological sample using a specific antibody. The methods generally comprise contacting the sample with a specific antibody and detecting binding between the antibody and molecules of the sample. Specific antibody binding, when compared to a suitable control, is an indication that a polypeptide of interest is present in the sample. Suitable controls include a sample known not to contain the polypeptide, and a sample contacted with a non-specific antibody, e.g., an anti-idiotype antibody.

[0477] A variety of methods to detect specific antibody-antigen interactions are known in the art, e.g., standard immunohistological methods, immunoprecipitation, enzyme immunoassay, and radioimmunoassay. The specific antibody can be detectably labeled, either directly or indirectly, as described at length herein, and cells are permeabilized to stain cytoplasmic molecules. Briefly, antibodies are added to a cell sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody may be labeled with radioisotopes, enzymes, fluorescers, chemiluminescers, or other labels for direct detection.. Alternatively, specific-binding pairs may be used, involving, e.g., a second stage antibody or reagent that is detectably-labeled, as described above. Such reagents and their methods of use are well known in the art

[0478] Alternatively, a biological sample can be brought into contact with an immobilized antibody on a solid support or carrier, such as nitrocellulose, that is capable of immobilizing cells, cell particles, or soluble proteins. The antibody can be attached (coupled) to an insoluble support, such as a polystyrene plate or a bead. After contacting the sample, the support can then be washed with suitable buffers, followed by contacting with a detectably-labeled specific antibody. Detection methods are known in the art and will be chosen as appropriate to the signal emitted by the detectable label. Detection is generally accomplished in comparison to suitable controls, and to appropriate standards.

[0479] The present invention further provides methods for detecting the presence and/or levels of enzymatic activity of a subject polypeptide in a biological

sample. The methods generally involve contacting the sample with a substrate that yields a detectable product upon being acted upon by a subject polypeptide, and detecting a product of the enzymatic reaction. Further, polypeptides that are subsets of the complete sequences of the subject proteins may be used to identify and investigate parts of the protein important for function.

[0480] The present invention further includes methods for monitoring activity of a polypeptide through observation of phenotypic changes in a cell containing such polypeptide, such as growth or differentiation, or the ability of such a cell to secrete a molecule that can be detected, such as through chemical methods or through its effect on another cell, such as cell activation.

5. Modulating mRNA and Peptides in Biological Samples

[0481] The present invention provides screening methods for identifying agents that modulate the level of a mRNA molecule of the invention, agents that modulate the level of a polypeptide of the invention, and agents that modulate the biological activity of a polypeptide of the invention. In some embodiments, the assay is cell-free; in others, it is cell-based. Where the screening assay is a binding assay, one or more of the molecules can be joined to a label, where the label can directly or indirectly provide a detectable signal.

[0482] As discussed above, the invention encompasses endogenous polynucleotides of the invention that encode mRNA and/or polypeptides of interest. Again as discussed previously, the invention also encompasses exogenous polynucleotides that encode mRNA or polypeptides of the invention. For example, the polynucleotide can reside within a recombinant vector which is introduced into the cell. For example, a recombinant vector can comprise an isolated transcriptional regulatory sequence which is associated in nature with a nucleic acid, such as a promoter sequence operably linked to sequences coding for a polypeptide of the invention; or the transcriptional control sequences can be operably linked to coding sequences for a polypeptide fusion protein comprising a polypeptide of the invention fused to a polypeptide that facilitates detection.

[0483] In these embodiments, the candidate agent is combined with a cell possessing a polynucleotide transcriptional regulatory element operably linked to a polypeptide-coding sequence of interest, e.g., a subject cDNA or its genomic component; and determining the agent's effect on polynucleotide expression, as measured, for example by the level of mRNA, polypeptide, or fusion polypeptide

[0484]   In other embodiments, for example, a recombinant vector can comprise an isolated polynucleotide transcriptional regulatory sequence, such as a promoter sequence, operably linked to a reporter gene (e.g., β-galactosidase, CAT, luciferase, or other gene that can be easily assayed for expression). In these embodiments, the method for identifying an agent that modulates a level of expression of a polynucleotide in a cell comprises combining a candidate agent with a cell comprising a transcriptional regulatory element operably linked to a reporter gene; and determining the effect of said agent on reporter gene expression.

[0485]   Known methods of measuring mRNA levels can be used to identify agents that modulate mRNA levels, including, but not limited to, PCR with detectably-labeled primers. Similarly, agents that modulate polypeptide levels can be identified using standard methods for determining polypeptide levels, including, but not limited to an immunoassay such as ELISA with detectably-labeled antibodies.

[0486]   A wide variety of cell-based assays can also be used to identify agents that modulate eukaryotic or prokaryotic mRNA and/or polypeptide levels. Examples include transformed cells that over-express a cDNA construct and cells transformed with a polynucleotide of interest associated with an endogenously-associated promoter operably linked to a reporter gene. A control sample would comprise, for example, the same cell lacking the candidate agent. Expression levels are measured and compared in the test and control samples.

[0487]   The cells used in the assay are usually mammalian cells, including, but not limited to, rodent cells and human cells. The cells can be primary cell cultures or can be immortalized cell lines. Cell-based assays generally comprise the steps of contacting the cell with a test agent, forming a test sample, and, after a suitable time, assessing the agent's effect on macromolecule expression. That is, the mammalian cell line is transformed or transfected with a construct that results in expression of the polynucleotide, the cell is contacted with a test agent, and then mRNA or polypeptide levels are detected and measured using conventional assays

[0488]   A suitable period of time for contacting the agent with the cell can be determined empirically, and is generally a time sufficient to allow entry of the agent into the cell and to allow the agent to have a measurable effect on subject mRNA and/or polypeptide levels. Generally, a suitable time is between about 10 minutes and about 24 hours, including about 1 to about 8 hours. Alternatively, incubation periods may be between about 0.1 and about 1 hour, selected for example for optimum

activity or to facilitate rapid high-throughput screening. Where the polypeptide is expressed on the cell surface, however, a shorter length of time may be sufficient. Incubations are performed at any suitable temperature, i.e., between about 4°C and about 40°C. The contact and incubation steps can be followed by a washing step to remove unbound components, i.e., a label that would give rise to a background signal during subsequent detection of specifically-bound complexes.

[0489] A variety of assay configurations and protocols are known in the art. For example, one of the components can be bound to a solid support, and the remaining components contacted with the support bound component. Remaining components may be added at different times or at substantially the same time. Further, where the interacting protein is a second subject protein, the effect of the test agent on binding can be determined by determining the effect on multimization of the subject protein.

[0490] The present invention further provides methods of identifying agents that modulate a biological activity of a polypeptide of the invention. The method generally comprises contacting a test agent with a sample containing a subject polypeptide and assaying a biological activity of the subject polypeptide in the presence of the test agent. An increase or a decrease in the assayed biological activity in comparison to the activity in a suitable control (e.g., a sample comprising a subject polypeptide in the absence of the test agent) is an indication that the substance modulates a biological activity of the subject polypeptide. The mixture of components is added in any order that provides for the requisite interaction..

[0491] External and internal processes that can affect modulation of a macromolecule of the invention include, but are not limited to, infection of a cell by a microorganism, including, but not limited to, a bacterium (e.g., *Mycobacterium* spp., *Shigella*, or *Chlamydia*), a protozoan (e.g., *Trypanosoma* spp., *Plasmodium* spp., or *Toxoplasma* spp.), a fungus, a yeast (e.g., *Candida* spp.), or a virus (including viruses that infect mammalian cells, such as human immunodeficiency virus, foot and mouth disease virus, Epstein-Barr virus, and viruses that infect plant cells); change in pH of the medium in which a cell is maintained or a change in internal pH; excessive heat relative to the normal range for the cell or the multicellular organism; excessive cold relative to the normal range for the cell or the multicellular organism; an effector molecule such as a hormone, a cytokine, a chemokine, a neurotransmitter; an ingested or applied drug; a ligand for a cell-surface receptor; a ligand for a receptor that exists

internally in a cell, e.g., a nuclear receptor; hypoxia; light; dark; sleep patterns; electrical charge; ion concentration of the medium in which a cell is maintained or an internal ion concentration, exemplary ions including sodium ions, potassium ions, chloride ions, calcium ions, and the like; presence or absence of a nutrient; metal ions; a transcription factor; mitogens, including, but not limited to, lipopolysaccharide (LPS), pokeweed mitogen; antigens; a tumor suppressor; and cell-cell contact and must be taken into consideration in the screening assay.

[0492] A variety of other reagents can be included in the screening assay. These include salts, neutral proteins, e.g., albumin, detergents, and other compounds that facilitate optimal binding and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, or anti-microbial agents, etc., can be used.

[0493] Accordingly, the present invention provides a method for identifying an agent, particularly a biologically active agent that modulates the level of expression of a nucleic acid in a cell, the method comprising: combining a candidate agent to be tested with a cell comprising a nucleic acid that encodes a polypeptide, and determining the agent's effect on polypeptide expression.

[0494] Some embodiments will detect agents that decrease the biological activity of a molecule of the invention. Maximal inhibition of the activity is not always necessary, or even desired, in every instance to achieve a therapeutic effect. Agents that decrease a biological activity can find use in treating disorders associated with the biological activity of the molecule. Alternatively, some embodiments will detect agents that increase a biological activity. Agents that increase a biological activity of a molecule of the invention can find use in treating disorders associated with a deficiency in the biological activity. Agents that increase or decrease a biological activity of a molecule of the invention can be selected for further study, and assessed for physiological attributes, i.e., cellular availability, cytotoxicity, or biocompatibility, and optimized as required. For example, a candidate agent is assessed for any cytotoxic activity it may exhibit toward the cell used in the assay using well-known assays, such as trypan blue dye exclusion, an MTT ([3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2 H-tetrazolium bromide]) assay, and the like.

[0495] A variety of different candidate agents can be screened by the above methods. Candidate agents encompass numerous chemical classes, as described above.

[0496] Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. Numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. For example, random peptide libraries obtained by yeast two-hybrid screens (Xu et al., 1997), phage libraries (Hoogenboom et al., 1998), or chemically generated libraries. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced, including antibodies produced upon immunization of an animal with subject polypeptides, or fragments thereof, or with the encoding polynucleotides. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and can be used to produce combinatorial libraries. Further, known pharmacological agents can be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, and amidification, etc, to produce structural analogs.

6. Kits

[0497] The present invention provides methods for diagnosing disease states based on the detected presence and/or level of polynucleotide or polypeptide in a biological sample, and/or the detected presence and/or level of biological activity of the polynucleotide or polypeptide. These detection methods can be provided as part of a kit. Thus, the invention further provides kits for detecting the presence and/or a level of a polynucleotide or polypeptide in a biological sample and/or or the detected presence and/or level of biological activity of the polynucleotide or polypeptide. Procedures using these kits can be performed by clinical laboratories, experimental laboratories, medical practitioners, or private individuals.

[0498] The kits of the invention will comprise a molecule of the invention. The kits for detecting a polynucleotide will also comprise a moiety that specifically hybridizes to a polynucleotide of the invention. The polynucleotide molecule can be of any length. For example, it can comprise a polynucleotide of at least 6, at least 7, at least 8, or at least 9 contiguous nucleotides of a molecule of the invention. Kits of the invention for detecting a subject polypeptide will comprise a moiety that specifically binds to a polypeptide of the invention; the moiety includes, but is not limited to, a polypeptide-specific antibody.

[0499] The kits are useful in diagnostic applications. For example, the kit is useful to determine whether a given DNA sample isolated from an individual comprises an expressed nucleic acid, a polymorphism, or other variant.

[0500] Kits for detecting polynucleotides comprise a pair of nucleic acids in a suitable storage medium, e.g., a buffered solution, in a suitable container. The pair of isolated nucleic acid molecules serve as primers in an amplification reaction (e.g., a polymerase chain reaction). The kit can further include additional buffers, reagents for polymerase chain reaction (e.g., deoxynucleotide triphosphates (dNTP), a thermostable DNA polymerase, a solution containing Mg^{2+} ions (e.g., $MgCl_2$), and other components well known to those skilled in the art for carrying out a polymerase chain reaction). The kit can further include instructions for use, which may be provided in a variety of forms, e.g., printed information, or compact disc, and the like. The kit may further include reagents necessary to extract DNA from a biological sample and reagents for generating a cDNA copy of an mRNA. The kit may optionally provide additional useful components, including, but not limited to, buffers, developing reagents, labels, reacting surfaces, means for detections, control samples, standards, and interpretive information.

[0501] In some embodiments, a kit of the invention for detecting a polynucleotide, such as an mRNA encoding a polypeptide, comprises a pair of nucleic acids that function as "forward" and "reverse" primers that specifically amplify a cDNA copy of the mRNA. The "forward" and "reverse" primers are provided as a pair of isolated nucleic acid molecules, each from about 10 to about 200 nucleotides in length, the first nucleic acid molecule of the pair comprising a sequence of at least about 10 contiguous nucleotides having 100% sequence identity to a nucleic acid sequence shown in from SEQ ID NOS.: 1 - 1231 and 2463 - 3697, and the second nucleic acid molecule of the pair comprising a sequence of at least about 10 contiguous nucleotides having 100% sequence identity to the reverse complement of a nucleic acid sequence shown in SEQ ID NOS.: 1 - 1231 and 2463 - 3697 , wherein the sequence of the second nucleic acid molecule is located 3' of the nucleic acid sequence of the first nucleic acid molecule. The primer nucleic acids are prepared using any known method, e.g., automated synthesis. In some embodiments, one or both members of the pair of nucleic acid molecules comprise a detectable label.

[0502] Where the kit provides for polypeptide detection, it can include one or more specific antibodies. In some embodiments, the antibody specific to the

144

polypeptide is detectably labeled. In other embodiments, the antibody specific to the polypeptide is not labeled; instead, a second, detectably-labeled antibody is provided that binds to the specific antibody. The kit may further include blocking reagents, buffers, and reagents for developing and/or detecting the detectable marker. The kit may further include instructions for use, controls, and interpretive information.

[0503] Where the kit provides for detecting enzymatic activity, it includes a substrate that provides for a detectable product when acted upon by a polypeptide of interest. The kit may further include reagents necessary to detect and develop the detectable marker.

[0504] The present invention provides for kits with unit doses of an active agent. These agents are described in more detail below. In some embodiments, the agent is provided in oral or injectable doses. Such kits will comprise containers containing the unit doses and an informational package insert describing the use and attendant benefits of the drugs in treating a condition of interest.

Therapeutic Compositions

[0505] The invention further provides agents identified using a screening assay of the invention, and compositions comprising the agents, subject polypeptides, subject polynucleotides, recombinant vectors, and/or host cells, including pharmaceutical compositions for therapeutic administration. The subject compositions can be formulated using well-known reagents and methods. These compositions can include a buffer, which is selected according to the desired use of the agent, polypeptide, polynucleotide, recombinant vector, or host cell, and can also include other substances appropriate to the intended use. Those skilled in the art can readily select an appropriate buffer, a wide variety of which are known in the art, suitable for an intended use.

1. Excipients and Formulations

[0506] In some embodiments, compositions are provided in formulation with pharmaceutically acceptable excipients, a wide variety of which are known in the art (Gennaro, 2000; Ansel et al., 1999; Kibbe et al., 2000). Pharmaceutically acceptable excipients, such as vehicles, adjuvants, carriers or diluents, are readily available to the public. Moreover, pharmaceutically acceptable auxiliary substances, such as pH adjusting and buffering agents, tonicity adjusting agents, stabilizers, wetting agents and the like, are readily available to the public.

145

[0507] In pharmaceutical dosage forms, the compositions of the invention can be administered in the form of their pharmaceutically acceptable salts, or they can also be used alone or in appropriate association, as well as in combination, with other pharmaceutically active compounds. The subject compositions are formulated in accordance to the mode of potential administration. Administration of the agents can be achieved in various ways, including oral, buccal, nasal, rectal, parenteral, intraperitoneal, intradermal, transdermal, subcutaneous, intravenous, intra-arterial, intracardiac, intraventricular, intracranial, intratracheal, and intrathecal administration, etc., or otherwise by implantation or inhalation. Thus, the subject compositions can be formulated into preparations in solid, semi-solid, liquid or gaseous forms, such as tablets, capsules, powders, granules, ointments, solutions, suppositories, injections, inhalants and aerosols. The following methods and excipients are merely exemplary and are in no way limiting.

[0508] For oral preparations, the agents, polynucleotides, and polypeptides can be used alone or in combination with appropriate additives to make tablets, powders, granules or capsules, for example, with conventional additives, such as lactose, mannitol, corn starch, or potato starch; with binders, such as crystalline cellulose, cellulose derivatives, acacia, corn starch, or gelatins; with disintegrators, such as corn starch, potato starch, or sodium carboxymethylcellulose; with lubricants, such as talc or magnesium stearate; and if desired, with diluents, buffering agents, moistening agents, preservatives, and flavoring agents.

[0509] Suitable excipient vehicles are, for example, water, saline, dextrose, glycerol, ethanol, or the like, and combinations thereof. In addition, if desired, the vehicle can contain minor amounts of auxiliary substances such as wetting or emulsifying agents or pH buffering agents. Actual methods of preparing such dosage forms are known, or will be apparent, to those skilled in the art (Remington, 1985). The composition or formulation to be administered will, in any event, contain a quantity of the agent adequate to achieve the desired state in the subject being treated.

[0510] The agents, polynucleotides, and polypeptides can be formulated into preparations for injection by dissolving, suspending or emulsifying them in an aqueous or nonaqueous solvent, such as vegetable or other similar oils, synthetic aliphatic acid glycerides, esters of higher aliphatic acids or propylene glycol; and if desired, with conventional additives such as solubilizers, isotonic agents, suspending

agents, emulsifying agents, stabilizers and preservatives. Other formulations for oral or parenteral delivery can also be used, as conventional in the art

[0511] The agents, polynucleotides, and polypeptides can be utilized in aerosol formulation to be administered via inhalation. The compounds of the present invention can be formulated into pressurized acceptable propellants such as dichlorodifluoromethane, propane, nitrogen, and the like. Further, the agent, polynucleotides, or polypeptide composition may be converted to powder form for administration intranasally or by inhalation, as conventional in the art.

[0512] Furthermore, the agents can be made into suppositories by mixing with a variety of bases such as emulsifying bases or water-soluble bases. The compounds of the present invention can be administered rectally via a suppository. The suppository can include vehicles such as cocoa butter, carbowaxes and polyethylene glycols, which melt at body temperature, yet are solidified at room temperature.

[0513] A polynucleotide, polypeptide, or other modulator, can also be introduced into tissues or host cells by other routes, such as viral infection, microinjection, or vesicle fusion. For example, expression vectors can be used to introduce nucleic acid compositions into a cell as described above. Further, jet injection can be used for intramuscular administration (Furth et al., 1992). The DNA can be coated onto gold microparticles, and delivered intradermally by a particle bombardment device, or "gene gun" as described in the literature (Tang et al., 1992), where gold microprojectiles are coated with the DNA, then bombarded into skin cells.

[0514] Unit dosage forms for oral or rectal administration such as syrups, elixirs, and suspensions can be provided wherein each dosage unit, for example, teaspoonful, tablespoonful, tablet, or suppository, contains a predetermined amount of the composition containing one or more agents. Similarly, unit dosage forms for injection or intravenous administration can comprise the agent(s) in a composition as a solution in sterile water, normal saline or another pharmaceutically acceptable carrier.

2. Active Agents (or Modulators)

[0515] The nucleic acid, polypeptide, and modulator compositions of the subject invention find use as therapeutic agents in situations where one wishes to modulate an activity of a subject polypeptide in a host, particularly the activity of the subject polypeptides, or to provide or inhibit the activity at a particular anatomical

site. Thus, the compositions are useful in treating disorders associated with an activity of a subject polypeptide. The following provides further details of active agents of the present invention.

a) Antisense Oligonucleotides

[0516]   In certain embodiments of the invention, the active agent is an agent that modulates, and generally decreases or down regulates, the expression of a gene encoding a target protein in a host, i.e., antisense molecules. Anti-sense reagents include antisense oligonucleotides (ODN), i.e., synthetic ODN having chemical modifications from native nucleic acids, or nucleic acid constructs that express such anti-sense molecules as RNA. The antisense sequence is complementary to the mRNA of the targeted gene, and inhibits expression of the targeted gene products. Antisense molecules inhibit gene expression through various mechanisms, e.g., by reducing the amount of mRNA available for translation, through activation of RNase H, or steric hindrance. One or a combination of antisense molecules can be administered, where a combination can comprise multiple different sequences.

[0517]   Antisense molecules can be produced by expression of all or a part of the target gene sequence in an appropriate vector, where the transcriptional initiation is oriented such that an antisense strand is produced as an RNA molecule. Alternatively, the antisense molecule is a synthetic oligonucleotide. Antisense oligonucleotides can be chemically synthesized by methods known in the art (Wagner et al., 1993; Milligan et al., 1993) Oligonucleotides can be chemically modified from the native phosphodiester structure to increase their intracellular stability and binding affinity, for example, as described in detail above. Antisense oligonucleotides will generally be at least about 7, at least about 12, or at least about 20 nucleotides in length, and not more than about 500, not more than about 50, or not more than about 35 nucleotides in length, where the length is governed by efficiency of inhibition, and specificity, including absence of cross-reactivity, and the like. Short oligonucleotides, of from about 7 to about 8 bases in length, can be strong and selective inhibitors of gene expression (Wagner et al., 1996).

[0518]   A specific region or regions of the endogenous sense strand mRNA sequence is chosen to be complemented by the antisense sequence. Selection of a specific sequence for the oligonucleotide can use an empirical method, where several candidate sequences are assayed for inhibition of expression of the target gene in an *in*

vitro or animal model. A combination of sequences can also be used, where several regions of the mRNA sequence are selected for antisense complementation.

[0519] As an alternative to anti-sense inhibitors, catalytic nucleic acid compounds, e.g., ribozymes, or anti-sense conjugates can be used to inhibit gene expression. Ribozymes can be synthesized *in vitro* and administered to the patient, or can be encoded in an expression vector, from which the ribozyme is synthesized in the targeted cell (WO 9523225; Beigelman et al., 1995). Examples of oligonucleotides with catalytic activity are described in WO 9506764. Conjugates of anti-sense ODN with a metal complex, *e.g.,* terpyridyl Cu(II), capable of mediating mRNA hydrolysis are described in Bashkin *et al.,* 1995.

b) Interfering RNA

[0520] In some embodiments, the active agent is an interfering RNA (RNAi), including dsRNAi. RNA interference provides a method of silencing eukaryotic genes. Double stranded RNA can induce the homology-dependent degradation of its cognate mRNA in *C. elegans*, fungi, plants, *Drosophila*, and mammals (Gaudilliere et al., 2002). Use of RNAi to reduce a level of a particular mRNA and/or protein is based on the interfering properties of double-stranded RNA derived from the coding regions of a gene. The technique reduces the time between identifying an interesting gene sequence and understanding its function, and thus is an efficient high-throughput method for disrupting gene function (O'Neil, 2001). RNAi can also help identify the biochemical mode of action of a drug and to identify other genes encoding products that can respond or interact with specific compounds.

[0521] In one embodiment of the invention, complementary sense and antisense RNAs derived from a substantial portion of the subject polynucleotide are synthesized *in vitro*. The resulting sense and antisense RNAs are annealed in an injection buffer, and the double-stranded RNA injected or otherwise introduced into the subject, i.e., in food or by immersion in buffer containing the RNA (Gaudilliere et al., 2002; O'Neil et al., 2001; WO99/32619). In another embodiment, dsRNA derived from a gene of the present invention is generated *in vivo* by simultaneously expressing both sense and antisense RNA from appropriately positioned promoters operably linked to coding sequences in both sense and antisense orientations.

c) Peptides and Modified Peptides

[0522] In some embodiments of the present invention, the active agent is a peptide. Suitable peptides include peptides of from about 3 amino acids to about 50,

from about 5 to about 30, or from about 10 to about 25 amino acids in length. In some embodiments, a peptide has a sequence of from about 3 amino acids to about 50, from about 5 to about 30, or from about 10 to about 25 amino acids of corresponding naturally-occurring protein. In some embodiments, a peptide exhibits one or more of the following activities: inhibits binding of a subject polypeptide to an interacting protein or other molecule; inhibits subject polypeptide binding to a second polypeptide molecule; inhibits a signal transduction activity of a subject polypeptide; inhibits an enzymatic activity of a subject polypeptide; or inhibits a DNA binding activity of a subject polypeptide.

[0523] Peptides can include naturally-occurring and non-naturally occurring amino acids. Peptides can comprise D-amino acids, a combination of D- and L-amino acids, and various "designer" amino acids (e.g., β-methyl amino acids, Cα-methyl amino acids, and Nα-methyl amino acids, etc.) to convey special properties. Additionally, peptides can be cyclic. Peptides can include non-classical amino acids in order to introduce particular conformational motifs. Any known non-classical amino acid can be used. Non-classical amino acids include, but are not limited to, 1,2,3,4-tetrahydroisoquinoline-3-carboxylate; (2S,3S)-methylphenylalanine, (2S,3R)-methyl-phenylalanine, (2R,3S)-methyl-phenylalanine and (2R,3R)-methyl-phenylalanine; 2-aminotetrahydronaphthalene-2-carboxylic acid; hydroxy-1,2,3,4-tetrahydroisoquinoline-3-carboxylate; β-carboline (D and L); HIC (histidine isoquinoline carboxylic acid); and HIC (histidine cyclic urea). Amino acid analogs and peptidomimetics can be incorporated into a peptide to induce or favor specific secondary structures, including, but not limited to, LL-Acp (LL-3-amino-2-propenidone-6-carboxylic acid), a β-turn inducing dipeptide analog; β-sheet inducing analogs; β-turn inducing analogs; α-helix inducing analogs; γ-turn inducing analogs; Gly-Ala turn analogs; amide bond isostere; or tretrazol, and the like.

[0524] A peptide can be a depsipeptide, which can be linear or cyclic (Kuisle et al., 1999). Linear depsipeptides can comprise rings formed through S–S bridges, or through an hydroxy or a mercapto group of an hydroxy-, or mercapto-amino acid and the carboxyl group of another amino- or hydroxy-acid but do not comprise rings formed only through peptide or ester links derived from hydroxy carboxylic acids. Cyclic depsipeptides contain at least one ring formed only through peptide or ester links, derived from hydroxy carboxylic acids.

[0525] Peptides can be cyclic or bicyclic. For example, the C-terminal carboxyl group or a C-terminal ester can be induced to cyclize by internal displacement of the -OH or the ester (-OR) of the carboxyl group or ester respectively with the N-terminal amino group to form a cyclic peptide. For example, after synthesis and cleavage to give the peptide acid, the free acid is converted to an activated ester by an appropriate carboxyl group activator such as dicyclohexylcarbodiimide (DCC) in solution, for example, in methylene chloride (CH_2Cl_2), dimethyl formamide (DMF) mixtures. The cyclic peptide is then formed by internal displacement of the activated ester with the N-terminal amine. Internal cyclization as opposed to polymerization can be enhanced by use of very dilute solutions. Methods for making cyclic peptides are well known in the art.

[0526] The term "bicyclic" refers to a peptide with two ring closures formed by covalent linkages between amino acids. A covalent linkage between two nonadjacent amino acids constitutes a ring closure, as does a second covalent linkage between a pair of adjacent amino acids which are already linked by a covalent peptide linkage. The covalent linkages forming the ring closures can be amide linkages, i.e., the linkage formed between a free amino on one amino acid and a free carboxyl of a second amino acid, or linkages formed between the side chains or "R" groups of amino acids in the peptides. Thus, bicyclic peptides can be "true" bicyclic peptides, i.e., peptides cyclized by the formation of a peptide bond between the N-terminus and the C-terminus of the peptide, or they can be "depsi-bicyclic" peptides, i.e., peptides in which the terminal amino acids are covalently linked through their side chain moieties.

[0527] A desamino or descarboxy residue can be incorporated at the terminal ends of the peptide, so that there is no terminal amino or carboxyl group, to decrease susceptibility to proteases or to restrict conformation. C-terminal functional groups include amide, amide lower alkyl, amide di (lower alkyl), lower alkoxy, hydroxy, and carboxy, and the lower ester derivatives thereof, and the pharmaceutically acceptable salts thereof.

[0528] In addition to the foregoing N-terminal and C-terminal modifications, a peptide or peptidomimetic can be modified with or covalently coupled to one or more of a variety of hydrophilic polymers to increase solubility and circulation half-life of the peptide. Suitable nonproteinaceous hydrophilic polymers for coupling to a peptide include, but are not limited to, polyalkylethers as exemplified by polyethylene

glycol and polypropylene glycol, polylactic acid, polyglycolic acid, polyoxyalkenes, polyvinylalcohol, polyvinylpyrrolidone, cellulose and cellulose derivatives, dextran, and dextran derivatives. Generally, such hydrophilic polymers have an average molecular weight ranging from about 500 to about 100,000 daltons, from about 2,000 to about 40,000 daltons, or from about 5,000 to about 20,000 daltons. The peptide can be derivatized with or coupled to such polymers using any of the methods set forth in Zallipsky, 1995; Monfardini et al., 1995; U.S. Pat. Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192; 4,179,337, or WO 95/34326.

d) Antibodies

[0529] The invention provides antibodies that specifically recognize a particular polypeptide. Antibodies are obtained by immunizing a host animal with peptides, polynucleotides encoding polypeptides, or cells, each comprising all or a portion of the target protein ("immunogen"). Suitable host animals include rodents (e.g., mouse, rat, guinea pig, hamster), cattle (e.g., sheep, pig, cow, horse, goat), cat, dog, chicken, primate, monkey, and rabbit. The origin of the protein immunogen can be any species, including mouse, human, rat, monkey, avian, insect, reptile, or crustacean. The host animal will generally be a different species than the immunogen, e.g., a human protein used to immunize mice. Methods of antibody production are well known in the art (Howard and Bethell, 2000; Harlow et al., 1998; Harlow and Lane, 1988).

[0530] The immunogen can comprise the complete protein, or fragments and derivatives thereof, or proteins expressed on cell surfaces. Immunogens comprise all or a part of one of the subject proteins, where these amino acids contain post-translational modifications, such as glycosylation, found on the native target protein. Immunogens comprising protein extracellular domains are produced in a variety of ways known in the art, e.g., expression of cloned genes using conventional recombinant methods, or isolation from tumor cell culture supernatants, etc. The immunogen can also be expressed *in vivo* from a polynucleotide encoding the immunogenic peptide introduced into the host animal.

[0531] Polyclonal antibodies are prepared by conventional techniques. These include immunizing the host animal *in vivo* with the target protein (or immunogen) in substantially pure form, for example, comprising less than about 1% contaminant. The immunogen can comprise the complete target protein, fragments, or derivatives thereof. To increase the immune response of the host animal, the target

protein can be combined with an adjuvant; suitable adjuvants include alum, dextran, sulfate, large polymeric anions, and oil & water emulsions, e.g., Freund's adjuvant (complete or incomplete). The target protein can also be conjugated to synthetic carrier proteins or synthetic antigens. The target protein is administered to the host, usually intradermally, with an initial dosage followed by one or more, usually at least two, additional booster dosages. Following immunization, blood from the host will be collected, followed by separation of the serum from blood cells. The immunoglobulin present in the resultant antiserum can be further fractionated using known methods, such as ammonium salt fractionation, or DEAE chromatography and the like.

[0532] The method of producing polyclonal antibodies can be varied in some embodiments of the present invention. For example, instead of using a single substantially isolated polypeptide as an immunogen, one may inject a number of different immunogens into one animal for simultaneous production of a variety of antibodies. In addition to protein immunogens, the immunogens can be nucleic acids (e.g., in the form of plasmids or vectors) that encode the proteins, with facilitating agents, such as liposomes, microspheres, etc., or without such agents, such as "naked" DNA.

[0533] Antibodies can also be prepared using a library approach. Briefly, mRNA is extracted from the spleens of immunized animals to isolate antibody-encoding sequences. The extracted mRNA may be used to make cDNA libraries. Such a cDNA library may be normalized and subtracted in a manner conventional in the art, for example, to subtract out cDNA hybridizing to mRNA of non-immunized animals. The remaining cDNA may be used to create proteins and for selection of antibody molecules or fragments that specifically bind to the immunogen. The cDNA clones of interest, or fragments thereof, can be introduced into an *in vitro* expression system to produce the desired antibodies, as described herein.

[0534] In a further embodiment, polyclonal antibodies can be prepared using phage display libraries, conventional in the art. In this method, a collection of bacteriophages displaying antibody properties on their surfaces are made to contact subject polypeptides, or fragments thereof. Bacteriophages displaying antibody properties that specifically recognize the subject polypeptides are selected, amplified, for example, in *E. coli*, and harvested. Such a method typically produces single chain antibodies

[0535]   Monoclonal antibodies are also produced by conventional techniques, such as fusing an antibody-producing plasma cell with an immortal cell to produce hybridomas. Suitable animals will be used, e.g., to raise antibodies against a mouse polypeptide of the invention, the host animal will generally be a hamster, guinea pig, goat, chicken, or rabbit, and the like. Generally, the spleen and/or lymph nodes of an immunized host animal provide the source of plasma cells, which are immortalized by fusion with myeloma cells to produce hybridoma cells. Culture supernatants from individual hybridomas are screened using standard techniques to identify clones producing antibodies with the desired specificity. The antibody can be purified from the hybridoma cell supernatants or from ascites fluid present in the host by conventional techniques, e.g., affinity chromatography using antigen, e.g., the subject protein, bound to an insoluble support, i.e., protein A sepharose, etc.

[0536]   The antibody can be produced as a single chain, instead of the normal multimeric structure of the immunoglobulin molecule. Single chain antibodies have been previously described (i.e., Jost et al., 1994). DNA sequences encoding parts of the immunoglobulin, for example, the variable region of the heavy chain and the variable region of the light chain are ligated to a spacer, such as one encoding at least about four small neutral amino acids, i.e., glycine or serine. The protein encoded by this fusion allows the assembly of a functional variable region that retains the specificity and affinity of the original antibody.

[0537]   The invention also provides intrabodies that are intracellularly expressed single-chain antibody molecules designed to specifically bind and inactivate target molecules inside cells. Intrabodies have been used in cell assays and in whole organisms (Chen et al., 1994; Hassanzadeh et al., 1998). Inducible expression vectors can be constructed with intrabodies that react specifically with a protein of the invention. These vectors can be introduced into host cells and model organisms.

[0538]   The invention also provides "artificial" antibodies, e.g., antibodies and antibody fragments produced and selected in vitro. In some embodiments, these antibodies are displayed on the surface of a bacteriophage or other viral particle, as described above. In other embodiments, artificial antibodies are present as fusion proteins with a viral or bacteriophage structural protein, including, but not limited to, M13 gene III protein. Methods of producing such artificial antibodies are well known in the art (U.S. Patent Nos. 5,516,637; 5,223,409; 5,658,727; 5,667,988; 5,498,538;

154

5,403,484; 5,571,698; and 5,625,033). The artificial antibodies, selected for example, on the basis of phage binding to selected antigens, can be fused to a Fc fragment of an immunoglobulin for use as a therapeutic, as described, for example, in US 5,116,964 or WO 99/61630. Antibodies of the invention can be used to modulate biological activity of cells, either directly or indirectly. A subject antibody can modulate the activity of a target cell, with which it has primary interaction, or it can modulate the activity of other cells by exerting secondary effects, i.e., when the primary targets interact or communicate with other cells. The antibodies of the invention can be administered to mammals, and the present invention includes such administration, particularly for therapeutic and/or diagnostic purposes in humans.

[0539] Antibodies may be administered by injection systemically, such as by intravenous injection; or by injection or application to the relevant site, such as by direct injection into a tumor, or direct application to the site when the site is exposed in surgery; or by topical application, such as if the disorder is on the skin, for example.

[0540] For *in vivo* use, particularly for injection into humans, in some embodiments it is desirable to decrease the antigenicity of the antibody. An immune response of a recipient against the antibody may potentially decrease the period of time that the therapy is effective. Methods of humanizing antibodies are known in the art. The humanized antibody can be the product of an animal having transgenic human immunoglobulin genes, e.g., constant region genes (e.g., Grosveld and Kolias, 1992; Murphy and Carter, 1993; Pinkert, 1994; and International Patent Applications WO 90/10077 and WO 90/04036). Alternatively, the antibody of interest can be engineered by recombinant DNA techniques to substitute the CH1, CH2, CH3, hinge domains, and/or the framework domain with the corresponding human sequence (see, e.g., WO 92/02190). Both polyclonal and monoclonal antibodies made in non-human animals may be "humanized" before administration to human subjects.

[0541] Chimeric immunoglobulin genes constructed with immunoglobulin cDNA are known in the art (Liu et al. 1987a; Liu et al. 1987b). Messenger RNA is isolated from a hybridoma or other cell producing the antibody and used to produce cDNA. The cDNA of interest can be amplified by the polymerase chain reaction using specific primers (U.S. Patent nos. 4,683,195 and 4,683,202). Alternatively, a library is made and screened to isolate the sequence of interest. The DNA sequence encoding the variable region of the antibody is then fused to human constant region

sequences. The sequences of human constant regions genes are known in the art
(Kabat et al., 1991). Human C region genes are readily available from known clones.
The choice of isotype will be guided by the desired effector functions, such as
complement fixation, or antibody-dependent cellular cytotoxicity. IgG1, IgG3 and
IgG4 isotypes, and either of the kappa or lambda human light chain constant regions
can be used. The chimeric, humanized antibody is then expressed by conventional
methods.

[0542]   Consensus sequences of heavy ("H") and light ("L") J regions can be
used to design oligonucleotides for use as primers to introduce useful restriction sites
into the J region for subsequent linkage of V region segments to human C region
segments. C region cDNA can be modified by site directed mutagenesis to place a
restriction site at the analogous position in the human sequence.

[0543]   A convenient expression vector for producing antibodies is one that
encodes a functionally complete human CH or CL immunoglobulin sequence, with
appropriate restriction sites engineered so that any VH or VL sequence can be easily
inserted and expressed, such as plasmids, retroviruses, YACs, or EBV derived
episomes, and the like. In such vectors, splicing usually occurs between the splice
donor site in the inserted J region and the splice acceptor site preceding the human C
region, and also at the splice regions that occur within the human CH exons.
Polyadenylation and transcription termination occur at native chromosomal sites
downstream of the coding regions. The resulting chimeric antibody can be joined to
any strong promoter, including retroviral LTRs, e.g., SV-40 early promoter,
(Okayama, et al. 1983), Rous sarcoma virus LTR (Gorman et al. 1982), and Moloney
murine leukemia virus LTR (Grosschedl et al. 1985), or native immunoglobulin
promoters.

[0544]   In yet other embodiments, the antibodies can be fully human
antibodies. For example, xenogenic antibodies, which are produced in animals that
are transgenic for human antibody genes, can be employed. By xenogenic human
antibodies is meant antibodies that are fully human antibodies, with the exception that
they are produced in a non-human host that has been genetically engineered to
express human antibodies. (e.g., WO 98/50433; WO 98,24893 and WO 99/53049).

[0545]   Antibody fragments, such as Fv, F(ab')2 and Fab can be prepared by
cleavage of the intact protein, e.g., by protease or chemical cleavage. These
fragments can include heavy and light chain variable regions. Alternatively, a

truncated gene can be designed, e.g., a chimeric gene encoding a portion of the F(ab')$_2$ fragment that includes DNA sequences encoding the CH1 domain and hinge region of the H chain, followed by a translational stop codon. The antibodies of the present invention may be administered alone or in combination with other molecules for use as a therapeutic, for example, by linking the antibody to cytotoxic agent, as discussed above, or to a radioactive molecule. Radioactive antibodies that are specific to a cancer cell, disease cell, or virus-infected cell may be able to deliver a sufficient dose of radioactivity to kill such cancer cell, disease cell, or virus-infected cell. The antibodies of the present invention can also be used in assays for detection of the subject polypeptides. In some embodiments, the assay is a binding assay that detects binding of a polypeptide with an antibody specific for the polypeptide; the subject polypeptide or antibody can be immobilized, while the subject polypeptide and/or antibody can be detectably-labeled. For example, the antibody can be directly labeled or detected with a labeled secondary antibody. That is, suitable, detectable labels for antibodies include direct labels, which label the antibody to the protein of interest, and indirect labels, which label an antibody that recognizes the antibody to the protein of interest.

[0546] These labels include radioisotopes, including, but not limited to ^{64}Cu, ^{67}Cu, ^{90}Y, ^{124}I, ^{125}I, ^{131}I, ^{137}Cs, ^{186}Re, ^{211}At, ^{212}Bi, ^{213}Bi, ^{223}Ra, ^{241}Am, and ^{244}Cm; enzymes having detectable products (e.g., luciferase, β-galactosidase, and the like); fluorescers and fluorescent labels, e.g., as provided herein; fluorescence emitting metals, e.g., ^{152}Eu, or others of the lanthanide series, attached to the antibody through metal chelating groups such as EDTA; chemiluminescent compounds, e.g., luminol, isoluminol, or acridinium salts; and bioluminescent compounds, e.g., luciferin, or aequorin (green fluorescent protein), specific binding molecules, e.g., magnetic particles, microspheres, nanospheres, and the like.

[0547] Alternatively, specific-binding pairs may be used, involving, e.g., a second stage antibody or reagent that is detectably-labeled and that can amplify the signal. For example, a primary antibody can be conjugated to biotin, and horseradish peroxidase-conjugated strepavidin added as a second stage reagent. Digoxin and antidigoxin provide another such pair. In other embodiments, the secondary antibody can be conjugated to an enzyme such as peroxidase in combination with a substrate that undergoes a color change in the presence of the peroxidase. The absence or presence of antibody binding can be determined by various methods, including flow

cytometry of dissociated cells, microscopy, radiography, or scintillation counting. Such reagents and their methods of use are well known in the art.

e) Peptide Aptamers

[0548]   Another suitable agent for modulating an activity of a subject polypeptide is a peptide aptamer. Peptide aptamers are peptides or small polypeptides that act as dominant inhibitors of protein function. Peptide aptamers specifically bind to target proteins, blocking their functional ability (Kolonin and Finley, 1998). Due to the highly selective nature of peptide aptamers, they can be used not only to target a specific protein, but also to target specific functions of a given protein (e.g., a signaling function). Further, peptide aptamers can be expressed in a controlled fashion by use of promoters which regulate expression in a temporal, spatial or inducible manner. Peptide aptamers act dominantly, therefore, they can be used to analyze proteins for which loss-of-function mutants are not available.

[0549]   Peptide aptamers that bind with high affinity and specificity to a target protein can be isolated by a variety of techniques known in the art. Peptide aptamers can be isolated from random peptide libraries by yeast two-hybrid screens (Xu et al., 1997). They can also be isolated from phage libraries (Hoogenboom et al., 1998) or chemically generated peptides/libraries.

Therapeutic Applications: Methods of Use

[0550]   The instant invention provides various therapeutic methods. In some embodiments, methods of modulating, including increasing and inhibiting, a biological activity of a subject protein are provided. In some embodiments, methods of modulating an enzymatic activity of a subject protein are provided. In some embodiments, methods of increasing the level of enzymatically active subject protein are provided, while in some embodiments, methods of decreasing a level of enzymatically active subject protein are provided.

[0551]   In some embodiments, methods of modulating enzymatic activity of a subject protein are provided. In other embodiments, methods of modulating a signal transduction activity of a subject protein are provided. In further embodiments, methods of modulating interaction of a subject protein with another, interacting protein or other macromolecule (e.g., DNA, carbohydrate, lipid) are provided. In further embodiments, methods of modulating transport activity of a subject protein are provided. In further embodiments, methods of modulating phopholipase activity of a subject protein are provided. In further embodiments, methods of modulating

polymerase activity of a subject protein are provided. In further embodiments, methods of modulating nuclease activity of a subject protein are provided.

[0552]   As mentioned above, an effective amount of the active agent (e.g., small molecule, antibody specific for a subject polypeptide, a subject polypeptide, or a subject polynucleotide) is administered to the host, where "effective amount" means a dosage sufficient to produce a desired effect or result. In some embodiments, the desired result is at least a reduction in a given biological activity of a subject polypeptide as compared to a control, for example, a decreased level of enzymatically active subject protein in the individual, or in a localized anatomical site in the individual. In further embodiments, the desired result is at least an increase in a biological activity of a subject polypeptide as compared to a control, for example an increased level of enzymatically active subject protein in the individual, or in a localized anatomical site in the individual.

[0553]   Typically, the compositions of the instant invention will contain from less than about 1% to about 95% of the active ingredient, about 10% to about 50%. Generally, between about 100 mg and about 500 mg will be administered to a child and between about 500 mg and about 5 grams will be administered to an adult.

[0554]   Other effective dosages can be readily determined by one of ordinary skill in the art through routine trials establishing dose response curves, for example, the amount of agent necessary to increase a level of active subject polypeptide can be calculated from *in vitro* experimentation. Those of skill will readily appreciate that dose levels can vary as a function of the specific compound, the severity of the symptoms, and the susceptibility of the subject to side effects, and preferred dosages for a given compound are readily determinable by those of skill in the art by a variety of means. For example, in order to calculate the polypeptide, polynucleotide, or modulator dose, those skilled in the art can use readily available information with respect to the amount necessary to have the desired effect, depending upon the particular agent used.

[0555]   The active agent(s) can be administered to the host via any convenient means capable of resulting in the desired result. Administration is generally by injection and often by injection to a localized area. The frequency of administration will be determined by the care given based on patient responsiveness. For example, the agents may be administered daily, weekly, or as conventionally determined appropriate.

[0556] A variety of hosts are treatable according to the subject methods. The host, or patient, may be from any animal species, and will generally be mammalian, e.g., primate sp., e.g., monkeys, chimpanzees, and particularly humans; rodents, including mice, rats and hamsters, guinea pig; rabbits; cattle, including equines, bovines, pig, sheep, goat, canines; felines; etc. Animal models are of interest for experimental investigations, providing a model for treatment of human disease.

Proliferative Conditions

[0557] In some embodiments, a protein of the present invention is involved in the control of cell proliferation, and an agent of the invention inhibits undesirable cell proliferation. Such agents are useful for treating disorders that involve abnormal cell proliferation, including, but not limited to, cancer, psoriasis, and scleroderma. Whether a particular agent and/or therapeutic regimen of the invention is effective in reducing unwanted cellular proliferation, e.g., in the context of treating cancer, can be determined using standard methods. For example, the number of cancer cells in a biological sample (e.g., blood, a biopsy sample, and the like), can be determined. The tumor mass can be determined using standard radiological or biochemical methods.

[0558] Tumors that can be treated using the methods of the instant invention include carcinomas, e.g., colorectal, prostate, breast, bone, kidney, skin, melanoma, ductal, endometrial, stomach or other organ of the gastrointestinal tract, pancreatic, mesothelioma, dysplastic oral mucosa, invasive oral cancer, non-small cell lung carcinoma ("NSCL"), transitional and squamous cell urinary carcinoma; brain cancer and neurological malignancies, *e.g.,* neuroblastoma, glioblastoma, astrocytoma, and gliomas; lymphomas and leukemias such as myeloid leukemia, myelogenous leukemia, hematological malignancies, such as childhood acute leukemia, non-Hodgkin's lymphomas, chronic lymphocytic leukemia, malignant cutaneous T-cell lymphoma, mycosis fungoides, non-MF cutaneous T-cell lymphoma, lymphomatoid papulosis, T-cell rich cutaneous lymphoid hyperplasia, bullous pemphigoid, discoid lupus erythematosus, lichen planus, and human follicular lymphoma; cancers of the reproductive system, e.g., cervical and ovarian cancers and testicular cancers; liver cancers including hepatocellular carcinoma ("HCC") and tumors of the biliary duct; multiple myelomas; tumors of the esophageal tract; other lung cancers and tumors including small cell and clear cell; Hodgkin's lymphomas; adenocarcinoma; and sarcomas, including soft tissue sarcomas.

Immunotherapeutic Approaches to Proliferative Conditions

[0559] The polynucleotides, polypeptides, and modulators of the present invention find use in immunotherapy of hyperproliferative disorders, including cancer, neoplastic, and paraneoplastic disorders. That is, the subject molecules can correspond to tumor antigens, of which 1770 have been identified to date (Yu and Restifo, 2002). Immunotherapeutic approaches include passive immunotherapy and vaccine therapy and can accomplish both generic and antigen-specific cancer immunotherapy.

[0560] Passive immunity approaches involve antibodies of the invention that are directed toward specific tumor-associated antigens. Such antibodies can eradicate systemic tumors at multiple sites, without eradicating normal cells. In some embodiments, the antibodies are combined with radioactive components, as provided above, for example, combining the antibody's ability to specifically target tumors with the added lethality of the radioisotope to the tumor DNA.

[0561] Useful antibodies comprise a discrete epitope or a combination of nested epitopes, i.e., a 10-mer epitope and associated peptide multimers incorporating all potential 8-mers and 9-mers, or overlapping epitopes (Dutoit et al., 2002). Thus a single antibody can interact with one or more epitopes. Further, the antibody can be used alone or in combination with different antibodies, that all recognize either a single or multiple epitopes.

[0562] Neutralizing antibodies can provide therapy for cancer and proliferative disorders. Neutralizing antibodies that specifically recognize a secreted protein or peptide of the invention can bind to the secreted protein or peptide, e.g., in a bodily fluid or the extracellular space, thereby modulating the biological activity of the secreted protein or peptide. For example, neutralizing antibodies specific for secreted proteins or peptides that play a role in stimulating the growth of cancer cells can be useful in modulating the growth of cancer cells. Similarly, neutralizing antibodies specific for secreted proteins or peptides that play a role in the differentiation of cancer cells can be useful in modulating the differentiation of cancer cells.

[0563] Vaccine therapy involves the use of polynucleotides, polypeptides, or agents of the invention as immunogens for tumor antigens (Machiels et al., 2002). For example, peptide-based vaccines of the invention include unmodified subject polypeptides, fragments thereof, and MHC class I and class II-restricted peptide

(Knutson et al., 2001), comprising, for example, the disclosed sequences with universal, nonspecific MHC class II-restricted epitopes. Peptide-based vaccines comprising a tumor antigen can be given directly, either alone or in conjunction with other molecules. The vaccines can also be delivered orally by producing the antigens in transgenic plants that can be subsequently ingested (U.S. Patent No. 6,395,964).

[0564] In some embodiments, antibodies themselves can be used as antigens in anti-idiotype vaccines. That is, administering an antibody to a tumor antigen stimulates B cells to make antibodies to that antibody, which in turn recognize the tumor cells

[0565] Nucleic acid-based vaccines can deliver tumor antigens as polynucleotide constructs encoding the antigen. Vaccines comprising genetic material, such as DNA or RNA, can be given directly, either alone or in conjunction with other molecules. Administration of a vaccine expressing a molecule of the invention, e.g., as plasmid DNA, leads to persistent expression and release of the therapeutic immunogen over a period of time, helping to control unwanted tumor growth.

[0566] In some embodiments, nucleic acid-based vaccines encode subject antibodies. In such embodiments, the vaccines (e.g., DNA vaccines) can include post-transcriptional regulatory elements, such as the post-transcriptional regulatory acting RNA element (WPRE) derived from Woodchuck Hepatitis Virus. These post-transcriptional regulatory elements can be used to target the antibody, or a fusion protein comprising the antibody and a co-stimulatory molecule, to the tumor microenvironment (Pertl et al., 2003).

[0567] Besides stimulating anti-tumor immune responses by inducing humoral responses, vaccines of the invention can also induce cellular responses, including stimulating T-cells that recognize and kill tumor cells directly. For example, nucleotide-based vaccines of the invention encoding tumor antigens can be used to activate the CD8$^+$ cytotoxic T lymphocyte arm of the immune system.

[0568] In some embodiments, the vaccines activate T-cells directly, and in others they enlist antigen-presenting cells to activate T-cells. Killer T-cells are primed, in part, by interacting with antigen-presenting cells, i.e., dendritic cells. In some embodiments, plasmids comprising the nucleic acid molecules of the invention enter antigen-presenting cells, which in turn display the encoded tumor-antigens that

contribute to killer T-cell activation. Again, the tumor antigens can be delivered as plasmid DNA constructs, either alone or with other molecules.

[0569] In further embodiments, RNA can be used. For example, dendritic cells can be transfected with RNA encoding tumor antigens (Heiser et al., 2002; Mitchell and Nair, 2000). This approach overcomes the limitations of obtaining sufficient quantities of tumor material, extending therapy to patients otherwise excluded from clinical trials. For example, a subject RNA molecule isolated from tumors can be amplified using RT-PCR. In some embodiments, the RNA molecule of the invention is directly isolated from tumors and transfected into dendritic cells with no intervening cloning steps.

[0570] In some embodiments the molecules of the invention are altered such that the peptide antigens are more highly antigenic than in their native state. These embodiments address the need in the art to overcome the poor *in vivo* immunogenicity of most tumor antigens by enhancing tumor antigen immunogenicity via modification of epitope sequences (Yu and Restifo, 2002).

[0571] Another recognized problem of cancer vaccines is the presence of preexisting neutralizing antibodies. Some embodiments of the present invention overcome this problem by using viral vectors from non-mammalian natural hosts, i.e., avian pox viruses. Alternative embodiments that also circumvent preexisting neutralizing antibodies include genetically engineered influenza viruses, and the use of "naked" plasmid DNA vaccines that contain DNA with no associated protein. (Yu and Restifo, 2002).

[0572] All of the immunogenic methods of the invention can be used alone or in combination with other conventional or unconventional therapies. For example, immunogenic molecules can be combined with other molecules that have a variety of antiproliferative effects, or with additional substances that help stimulate the immune response, i.e., adjuvants or cytokines.

[0573] For example, in some embodiments, nucleic acid vaccines encode an alphaviral replicase enzyme, in addition to tumor antigens. This recently discovered approach to vaccine therapy successfully combines therapeutic antigen production with the induction of the apoptotic death of the tumor cell (Yu and Restifo, 2002).

[0574] In certain other embodiments, a DNA or RNA vaccine of the present invention can also be directed against the production of blood vessels in the vicinity of the tumor, a process called antiangiogenesis, thereby depriving the cancer cells of

nutrients. For example, the antiangiogenic molecules angiostatin (a fragment of plasminogen), endostatin (a fragment of collagen XVIII), interferon-γ, interferon-γ inducible protein 10, interleukin 12, thrombospondin, platelet factor-4, calreticulin, or its protein fragment vasostatin can be used to treat tumors by suppressing neovascularization and thereby inhibiting growth (Cheng et al., 2001). The antiangiogenesis approach can be used alone, or in conjunction with molecules directed to tumor antigens.

[0575] Furthermore, adjuvants can be used in conjunction with the antibodies and vaccines disclosed herein. Adjuvants help boost the general immune response, for example, concentrating immune cells to the specific area where they are needed. They can be added to a cancer vaccine itself or administered separately, and in some embodiments, a viral vector can be engineered to display adjuvant proteins on its surface.

[0576] Cytokines can also be used to help stimulate immune response. Cytokines act as chemical messengers, recruiting immune cells that help the killer T-cells to the site of attack. An example of a cytokine is granulocyte-macrophage colony-stimulating factor (GM-CSF), which stimulates the proliferation of antigen-presenting cells, thus boosting an organism's response to a cancer vaccine. As with adjuvants, cytokines can be used in conjunction with the antibodies and vaccines disclosed herein. For example, they can be incorporated into the antigen-encoding plasmid or introduced via a separate plasmid, and in some embodiments, a viral vector can be engineered to display cytokines on its surface.

Inflammation and Immunity

[0577] In other embodiments, e.g., where the subject polypeptide is involved in modulating inflammation or immune function, the invention provides agents for treating such inflammation or immune disorders. Disease states that are treatable using formulations of the invention include various types of arthritis such as rheumatoid arthritis and osteoarthritis, autoimmune thyroiditis, various chronic inflammatory conditions of the skin, such as psoriasis, the intestine, such as inflammatory bowel disease (IBD), insulin-dependent diabetes, autoimmune diseases such as multiple sclerosis (MS), intestinal immune disorders and systemic lupus erythematosis (SLE), allergic diseases, transplant rejections, adult respiratory distress syndrome, atherosclerosis, ischemic diseases due to closure of the peripheral

vasculature, cardiac vasculature, and vasculature in the central nervous system (CNS). After reading the present disclosure, those skilled in the art will recognize other disease states and/or symptoms which might be treated and/or mitigated by the administration of formulations of the present invention.

[0578] Neutralizing antibodies can provide immunosuppressive therapy for inflammatory and autoimmune disorders. Neutralizing antibodies can be used to treat disorders such as, for example, multiple sclerosis, rheumatoid arthritis, inflammatory bowel disease, transplant rejection, and psoriasis. Neutralizing antibodies that specifically recognize a secreted protein or peptide of the invention can bind to the secreted protein or peptide, e.g., in a bodily fluid or the extracellular space, thereby modulating the biological activity of the secreted protein or peptide. For example, neutralizing antibodies specific for secreted proteins or peptides that play a role in activating immune cells are useful as immunosuppressants.

Disorders Related to Cell Death

[0579] Where a polypeptide of the invention is involved in modulating cell death, an agent of the invention is useful for treating conditions or disorders relating to cell death (e.g., DNA damage, cell death, apoptosis). Cell death-related indications that can be treated using the methods of the invention to reduce cell death in a eukaryotic cell, include, but are not limited to, cell death associated with Alzheimer's disease, Parkinson's disease, rheumatoid arthritis, autoimmune thyroiditis, septic shock, sepsis, stroke, central nervous system inflammation, intestinal inflammation, osteoporosis, ischemia, reperfusion injury, cardiac muscle cell death associated with cardiovascular disease, polycystic kidney disease, cell death of endothelial cells in cardiovascular disease, degenerative liver disease, multiple sclerosis, amyotropic lateral sclerosis, cerebellar degeneration, ischemic injury, cerebral infarction, myocardial infarction, acquired immunodeficiency syndrome (AIDS), myelodysplastic syndromes, aplastic anemia, male pattern baldness, and head injury damage. Also included are conditions in which DNA damage to a cell is induced by external conditions, including but not limited to irradiation, radiomimetic drugs, hypoxic injury, chemical injury, and damage by free radicals. Also included are any hypoxic or anoxic conditions, e.g., conditions relating to or resulting from ischemia, myocardial infarction, cerebral infarction, stroke, bypass heart surgery, organ transplantation, and neuronal damage, etc.

[0580] DNA damage can be detected using any known method, including, but not limited to, a Comet assay (commercially available from Trevigen, Inc.), which is based on alkaline lysis of labile DNA at sites of damage; and immunological assays using antibodies specific for aberrant DNA structures, e.g., 8-OHdG.

[0581] Cell death can be measured using any known method, and is generally measured using any of a variety of known methods for measuring cell viability. Such assays are generally based on entry into the cell of a detectable compound (or a compound that becomes detectable upon interacting with, or being acted on by, an intracellular component) that would normally be excluded from a normal, living cell by its structurally and functionally intact cell membrane. Such compounds include substrates for intracellular enzymes, including, but not limited to, a fluorescent substrate for esterase; dyes that are excluded from living cells, including, but not limited to, trypan blue; and DNA-binding compounds, including, but not limited to, an ethidium compound such as ethidium bromide and ethidium homodimer, and propidium iodide.

[0582] Apoptosis, or programmed cell death, is a regulated process leading to cell death via a series of well-defined morphological changes. Programmed cell death provides a balance for cell growth and multiplication, eliminating unnecessary cells. The default state of the cell is to remain alive. A cell enters the apoptotic pathway when an essential factor is removed from the extracellular environment or when an internal signal is activated. Genes and proteins of the invention that suppress the growth of tumors by activating cell death provide the basis for treatment strategies for hyperproliferative disorders and conditions.

[0583] Apoptosis can be assayed using any known method. Assays can be conducted on cell populations or an individual cell, and include morphological assays and biochemical assays. A non-limiting example of a method of determining the level of apoptosis in a cell population is TUNEL (TdT-mediated dUTP nick-end labeling) labeling of the 3'-OH free end of DNA fragments produced during apoptosis (Gavrieli et al., 1992). The TUNEL method consists of catalytically adding a nucleotide, which has been conjugated to a chromogen system, a fluorescent tag, or the 3'-OH end of the 180-bp (base pair) oligomer DNA fragments, in order to detect the fragments. The presence of a DNA ladder of 180-bp oligomers is indicative of apoptosis. Procedures to detect cell death based on the TUNEL method are available

commercially, e.g., from Boehringer Mannheim (Cell Death Kit) and Oncor (Apoptag Plus).

[0584] Another marker that is currently available is annexin, sold under the trademark APOPTEST™. This marker is used in the "Apoptosis Detection Kit," which is also commercially available, e.g., from R&D Systems. During apoptosis, a cell membrane's phospholipid asymmetry changes such that the phospholipids are exposed on the outer membrane. Annexins are a homologous group of proteins that bind phospholipids in the presence of calcium. A second reagent, propidium iodide (PI), is a DNA binding fluorochrome. When a cell population is exposed to both reagents, apoptotic cells stain positive for annexin and negative for PI, necrotic cells stain positive for both, live cells stain negative for both. Other methods of testing for apoptosis are known in the art and can be used, including, e.g., the method disclosed in U.S. Patent No. 6,048,703.

Other Pathological Conditions

[0585] Other pathological conditions that can be treated using the methods of the instant invention include disorders of hematopoeisis, cell differentiation, disorders of ion channels, e.g., cystic fibrosis, and tissue or organ hypertrophy; viral disorders, including acquired immunodeficiency syndrome (AIDS), angiogenesis, metastasis, metabolic disorders such as diabetes and obesity, cardiovascular disorders such as congestive heart failure and stroke, male erectile dysfunction, and the disorders described throughout the specification.

Investigative Applications

[0586] The subject nucleic acid compositions find use in a variety of different investigative applications. Applications of interest include identifying genomic DNA sequence using molecules of the invention, identifying homologs of molecules of the invention, creating a source of novel promoter elements, identifying expression regulatory factors, creating a source of probes and primers for hybridization applications, identifying expression patterns in biological specimens; preparing cell or animal models to investigate the function of the molecules of the invention, and preparing *in vitro* models to investigate the function of the molecules of the invention.

Genomic DNA Sequences

[0587] Human genomic polynucleotide sequences corresponding to molecules of the present invention are identified by conventional means, such as, for

167

example, by probing a genomic DNA library with all or a portion of the polynucleotide sequences.

Homologs

[0588] Homologs are identified by any of a number of methods. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes, as described in detail above. Briefly, a fragment of the provided cDNA can be used as a hybridization probe against a cDNA library from the target organism of interest, under various stringency conditions, e.g., low stringency conditions. The probe can be a large fragment, or one or more short degenerate primers, and is typically labeled. Sequence identity can be determined by hybridization under stringent conditions, as described in detail above. Nucleic acids having a region of substantial identity or sequence similarity to the provided nucleic acid sequences, for example allelic variants, related genes, or genetically altered versions of the gene, bind to the provided sequences under less stringent hybridization conditions.

Promoter Elements and Expression Regulatory Factors

[0589] The sequence of the 5' flanking region can be utilized as promoter elements, including enhancer binding sites that provide for tissue-specific expression and developmental regulation in tissues where the subject genes are expressed, providing promoters that mimic the native pattern of expression. Naturally occurring polymorphisms in the promoter region are useful for determining natural variations in expression, particularly those that may be associated with disease. Promoters or enhancers that regulate the transcription of the polynucleotides of the present invention are obtainable by use of PCR techniques using human tissues, and one or more of the present primers.

[0590] Alternatively, mutations can be introduced into the promoter region to determine the effect of altering expression in experimentally defined systems. Methods for the identification of specific DNA motifs involved in the binding of transcriptional factors are known in the art, for example sequence similarity to known binding motifs, and gel retardation studies (Blackwell et al., 1995; Mortlock et al., 1996; Joulin and Richard-Foy, 1995).

[0591] The regulatory sequences can be used to identify *cis* acting sequences required for transcriptional or translational regulation of expression, especially in different tissues or stages of development, and to identify *cis* acting sequences and *trans*-acting factors that regulate or mediate expression. Such transcription or

translational control regions can be operably linked to a gene in order to promote expression of wild type genes or of proteins of interest in cultured cells, embryonic, fetal or adult tissues, and for gene therapy (Hooper, 1993).

Primers and Probes

[0592]   Small DNA fragments are useful as primers for reactions that involve nucleic acid hybridization, as described in detail above. Briefly, pairs of primers will be used in amplification reactions, such as PCR. Amplification primers hybridize to complementary strands of DNA, for example, under stringent conditions, and will prime towards each other. In some embodiments a pair of primers will generate an amplification product of at least about 50 nt, or at least about 100 nt. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages.

[0593]   The nucleotides can also be used as probes to identify genomic DNA or gene expression in a biological specimen, as described above and as is well established in the art. Briefly, DNA or mRNA is isolated from a cell sample. Detection of mRNA hybridizing to the subject sequence is indicative of gene expression in the sample. The mRNA can be amplified by RT-PCR, using reverse transcriptase to form a complementary DNA strand, followed by polymerase chain reaction amplification using primers specific for the subject DNA sequences. Alternatively, the mRNA sample is separated by gel electrophoresis, transferred to a suitable support, e.g., nitrocellulose, nylon, etc., and then probed with a fragment of the subject nucleotides as a probe. Other techniques, such as oligonucleotide ligation assays, in situ hybridizations, and hybridization to probes arrayed on a solid chip may also find use.

Targeted Mutations for *In Vivo* and *In Vitro* Models

[0594]   The sequence of a gene according to the subject invention, including flanking promoter regions and coding regions, can be mutated in various ways known in the art to generate targeted changes, i.e., changes in promoter strength, or sequence of the encoded protein, etc. The DNA sequence or protein product of such a mutation will usually be substantially similar to the sequences provided herein. The sequence changes can be substitutions, insertions, deletions, or a combination thereof. Deletions can further include larger changes, such as deletions of a domain or exon.

[0595]   Techniques for *in vitro* mutagenesis of cloned genes are known. Examples of protocols for site specific mutagenesis may be found in Gustin et al.,

1993; Barany 1985; Colicelli et al., 1985; Prentki et al., 1984. Methods for site specific mutagenesis can be found in Sambrook et al., 1989 (pp. 15.3-15.108); Weiner et al., 1993; Sayers et al. 1992; Jones and Winistorfer; Barton et al., 1990; Marotti and Tomich 1989; and Zhu, 1989. Such mutated genes can be used to study structure-function relationships of the subject proteins, or to alter properties of the protein that affect its function or regulation. Other modifications of interest include epitope tagging, e.g., with hemagglutinin (HA), FLAG, or c-*myc*. For studies of subcellular localization, fluorescent fusion proteins can be used.

[0596] The subject nucleic acids can be used to generate transgenic, non-human animals and/or site-specific gene modifications in cell lines; suitable methods are known in the art (Grosveld and Kollias, 1992; Hooper, 1993; Murphy and Carter, 1993; Pinkert, 1994). Thus, in some embodiments, the invention provides a non-human transgenic animal comprising, as a transgene integrated into the genome of the animal, a nucleic acid molecule comprising a sequence encoding a subject polypeptide in operable linkage with a promoter, such that the subject polypeptide-encoding nucleic acid molecule is expressed in a cell of the animal. Either a complete or partial sequence of a gene native to the host can be introduced. Alternatively, a complete or partial sequence of a gene exogenous to the host animal, e.g., a human sequence of the subject invention, can be introduced. Transgenic animals can be made through homologous recombination, where the endogenous locus is altered. Thus, DNA constructs for homologous recombination will comprise at least a portion of the human gene or of a gene native to the species of the host animal, wherein the gene has the desired genetic modification(s), and includes regions of homology to the target locus. Methods for generating mammalian cells having targeted gene modifications through homologous recombination are known in the art (Keown et al., 1990).

[0597] Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, and YACs. DNA constructs for random integration need not include regions of homology to mediate recombination.

[0598] Conveniently, markers for positive and negative selection are included. A detectable marker, such as *lac Z* can be introduced into a locus at which up-regulation of expression will result in a detectable change in phenotype.

[0599] Transformed ES or embryonic cells can be used to produce transgenic animals. An embryonic stem (ES) cell line can be a source of embryonic stem cells, or they can be newly obtained from a host animal, e.g., a mouse, rat, or guinea pig. The cells are grown on an appropriate fibroblast-feeder layer or in the presence of leukemia inhibiting factor (LIF). Following transformation, the cells are plated for growth onto a feeder layer in an appropriate medium. Cells containing the relevant construct can be detected by employing a selective medium and analyzing them for the occurrence of homologous recombination or integration of the construct. Positive colonies can be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old super-ovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant female animals that proceed to term. The resulting offspring are screened for the construct. By providing for a different phenotype of the blastocyst and the genetically modified cells, chimeric progeny can be readily detected.

[0600] The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs can be maintained as allogeneic or congenic grafts or transplants, or in *in vitro* culture. The transgenic animals can be any non-human mammal.

[0601] The modified cells or animals are useful in the study of gene function and regulation. For example, a series of small deletions and/or substitutions can be made in the host's native gene to determine the role of different exons in biological processes such as oncogenesis or signal transduction. Of interest is the use of genes to construct transgenic animal models for cancer, where expression of the subject protein is specifically reduced or absent. Specific constructs of interest include anti-sense constructs, which will block expression, expression of dominant negative mutations, and gene over-expression.

[0602] One can also provide for expression of the gene, e.g., a subject gene, or variants thereof, in cells or tissues where it is not normally expressed, at levels not normally present in such cells or tissues, or at abnormal times of development. One can also generate host cells (including host cells in transgenic animals) that comprise a heterologous nucleic acid molecule which encodes a polypeptide which functions to

modulate expression of an endogenous promoter or other transcriptional regulatory region, or the biological activity of a subject polypeptide.

[0603] The transgenic animals can also be used in functional studies, for example drug screening, to determine the effect of a candidate drug on a biological activity of a subject polypeptide.

Tables

Table 1. Characteristics of the Claimed Sequences, and of the Protein With the Highest Degree of Similarity to Each

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000214 N0_160000_ gene_predict ion1	142	0.81		gi|1082531|p ir||A55410	Ig light chain V region (variant CA2) - human (fragment)	
	HG1000323 N0_160000_ gene_predict ion1	527	0.75		gi|4557727|r ef|NP_00022 8.1|	lipoprotein lipase precursor [Homo sapiens]	
	HG1000323 N0_160000_ gene_predict ion2	222	0.28		gi|2144804|p ir||CGBO6C	collagen alpha 1(II) chain precursor - bovine (tentative sequence) (fragments)	
	HG1000327 N0_1000_ge ne_ prediction1	398	0.67	Secreted, lysosomal.	gi|27676136| ref|XP_2235 21.1|	similar to prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy);	Variant Gaucher disease and variant metachromatic leukodystrophy and other metabolic diseases; neuro:- metachromatic leuko- dystrophy variant (e.g., metachromatic leuko- dystrophy due to SAP-1 deficiency), neonatal hyperkinetic behavior (e.g., combined SAP deficiency), myoclonus (e.g., Combined

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000327 N0_160000_ gene_predict ion1	521	0.66	Secreted, lysosomal.	gi|27676136| ref|XP_2235 21.1|	similar to prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy);	respiratory insufficiency (e.g., combined SAP deficiency); GI: hepato-splenomegaly (e.g., combined SAP deficiency), splenomegaly (e.g., Gaucher disease, atypical); lab: prosaposin deficiency, marked glucosylceramide accumulation in the spleen (e.g., Gaucher disease, atypical); inheritance: autosomal dominant (10q21-q22). Variant Gaucher disease and variant metachromatic leukodystrophy and other metabolic diseases; neuro.: metachromatic leukodystrophy variant (e.g., metachromatic leukodystrophy due to SAP-1 deficiency), neonatal hyperkinetic behavior (e.g., combined SAP deficiency), myoclonus (e.g., Combined SAP deficiency); resp.:

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
							(e.g., combined SAP deficiency); GI: hepatosplenomegaly (e.g., combined SAP deficiency), splenomegaly (e.g., Gaucher disease, atypical); lab: prosaposin deficiency, marked glucosylceramide accumulation in the spleen (e.g., Gaucher disease, atypical); inheritance: autosomal dominant (10q21-q22).
	HG1000434 N0_160000_gene_predict ion1	108	0.12		gi|28917438| gb|EAA2714 0.1|	hypothetical protein [Neurospora crassa]	
	HG1000449 N0_160000_gene_predict ion1	85	0.89		gi|182204|gb |AAA52402. 1|	estrogen receptor	
	HG1000807 N0_160000_gene_predict ion1	392	0.39		gi|17384405| emb|CAD13 245.1| bA113O24.1	(similar to insulin-growth factor binding protein) [Homo sapiens]	
	HG1000807 N0_5000_ge ne_ prediction1	293	0.52		gi|17384405| emb|CAD13 245.1| bA113O24.1	(similar to insulin-growth factor binding protein) [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1001280 N0_160000_gene_predict ion1	2154	0.33		gi|665825|d bj|BAA1344 8.1|	Similar to Human C219-reactive pept (L34688) [Homo sapiens]	
	HG1000193 N0_160000_gene_predict ion1	118	0.5		gi|24797158| ref|NP_0058 99.2|	mannosidase, beta A, lysosomal; mannanase; mannase [Homo sapiens]	
	HG1000286 N0_160000_gene_predict ion1	360	0.88	Secreted.	gi|20141241| sp|P50454|C BP2_HUM AN	Collagen-binding protein 2 precursor (Colligin 2) (Rheumatoid arthritis related	therapeutic (angiogenesis inhibitor).
	HG1000992 N0_160000_gene_predict ion1	257	0.26		gi|27477294| ref|XP_2095 10.1|	similar to solute carrier family 9, member 7; nonselective sodium potassium/proton	
Kidney: Mahimkar et al., 2000.; Epididymis: Jury et al, 1999.	HG1001148 N0_160000_gene_predict ion1	317	0.82	Secreted, membrane anchored (Kratzschma r, et al., 1996).	gi|15778976| gb|AAH145 66.1|AAH14 566	a disintegrin and metalloproteinase domain 15 (metargidin) [Homo sapiens]	Anti-cancer target; anti kidney/renal.
	HG1001185 N0_160000_gene_predict ion2	112	0.58		gi|4502831|r ef|NP_00073 7.1|	cholinergic receptor, nicotinic, alpha polypeptide 7 precursor; a7 nicotinic	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
Cartilage; malignant melanoma.	HG1001280_NO_5000_gene_prediction1	118	0.74	Secreted.	gi|26351055|dbj|BAC391 64.1|	unnamed protein product [Mus musculus]	Tumor metastasis intervention; rheumatoid arthritis (RA) intervention, osteoarthritis (OA) intervention.
	HG1001302_NO_160000_gene_predict ion2	466	0.77		gi|14549163| sp|O00339| MTN2_HU MAN	Matrilin-2 precursor gi|1125762|gb|AAC5 1260.2| matrilin-2 precursor [Homo sapiens]	
	HG1000361_NO_160000_gene_predict ion1	1583	0.33		gi|24233577| ref|NP_7003 56.1|	hypothetical protein FLJ90440 [Homo sapiens]	
	HG1000361_NO_20000_g ene_predicti on1	315	0.78	Secreted.	gi|28503219| ref|XP_2829 50.1|	RIKEN cDNA 9430095K15 gene [Mus musculus]	
	HG1000792_NO_160000_gene_predict ion1	244	0.83		gi|27229118| ref|NP_0821 29.2|	RIKEN cDNA 0610006F02 [Mus musculus] gi|26324294|dbj|BAB 21956.2| unnamed protein product [Mus musculus]	Colon adenocarcinoma.
	HG1000934_NO_160000_gene_predict ion1	1583	0.33		gi|24233577| ref|NP_7003 56.1|	hypothetical protein FLJ90440 [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
Northern blot: liver, kidney, colon, small intestine, heart; Unigene CYP4F12: liver, spleen; colon; ascites; fetal eye; liver; hypothalamus; pooled colon, kidney, stomach; colon tumor; RER+; pooled human melanocyte, fetal heart, pregnant uterus; stomach; primary lung cystic fibrosis epithelial cells; dorsal root ganglia.	HG1000976 N0_160000_gene_predict ion1	389	0.85	Mitochondrial; microsomal.	gi|20532036|sp|Q9HCS2|CPFC_HUMAN	Cytochrome P450 4F12 (CYPIVF12) gi|25282619|pir||JC7594 cytochrome P450 enzyme, CYP4F12 isoform,	Cystic fibrosis; fertility.
	HG1000992 N0_10000_g ene_predicti on1	119	0.57		gi|27477294| ref|XP_2095 10.1|	similar to solute carrier family 9, member 7; nonselective sodium potassium/proton	
	HG1001185	65	1		gi|4502831|r	cholinergic receptc	

178

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_1000_gene_prediction1				ef|NP_000737.1|	nicotinic, alpha polypeptide 7 precursor; a7 nicotinic	
Brain, pancreas, prostate.	HG1001185 N0_160000_gene_prediction1	92	0.71	Integral membrane protein.	gi|4502831|ref|NP_000737.1|	cholinergic receptor, nicotinic, alpha polypeptide 7 precursor; a7 nicotinic	Neurological diseases; alcoholism.
	HG1001185 N0_1000_gene_prediction2	65	1		gi|4502831|ref|NP_000737.1|	cholinergic receptor, nicotinic, alpha polypeptide 7 precursor; a7 nicotinic	
Brain, pancreas, prostate.	HG1001185 N0_5000_gene_prediction1	92	0.71	Integral membrane protein	gi|4502831|ref|NP_000737.1|	cholinergic receptor, nicotinic, alpha polypeptide 7 precursor; a7 nicotinic	Neurological diseases; alcoholism.
	HG1001280 N0_10000_gene_prediction1	1159	0.57		gi|27477706|ref|XP_046126.4|	similar to Meningioma-expressed antigen 6/11 (MEA6) (MEA11) [Homo sapiens]	
	HG1000361 N0_10000_gene_prediction1	103	0.7	Secreted.	gi|28503219|ref|XP_282950.1|	RIKEN cDNA 9430095K15 gene [Mus musculus]	
	HG1001381 N0_1000_gene	68	0.47		gi|6090615|gb|AAF03259.1|	dihydropyridine receptor alpha 2 subunit [Homo	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	prediction1					sapiens]	mice; may be involved in immunomodulatory function.
	HG1000263 N0_5000_ge ne_ prediction1	253	0.26		gi|7512937|p ir||T08783	hypothetical protein DKFZp586O0120.1 - human (fragment)	
	HG1000579 N0_160000_ gene_predict ion1	368	0.93		gi|21619110| gb|AAH324 99.1|	Similar to Forssman glycolipid synthetase [Homo sapiens]	
	HG1000191 N0_160000_ gene_predict ion1	221	0.23		gi|27479383| ref|XP_2080 98.1|	similar to RE52392p [Drosophila melanogaster] [Homo sapiens]	
	HG1000296 N0_160000_ gene_predict ion2	165	0.6	Secreted.	gi|25054735| ref|XP_1928 39.1|	ATPas, class II, type 9B [Mus musculus]	Diagnostic marker.
	HG1000346 N0_1000_ge ne_ prediction1	59	0.98		gi|27754174| ref|NP_7761 69.1|	hypothetical protein MGC46680 [Homo sapiens]	
Widespread, except PBL.	HG1000963 N0_5000_ge ne_ prediction1	89	0.52	Endoplasmic reticulum (ER).	gi|14749486| ref|XP_0518 54.1|	similar to Mesoderm development candidate 2 [Homo sapiens]	Cancer; modulation of cell growth and differentiation, cell motility, invasion, metastasis.
	HG1000610 N0 160000	375	0.31		gi|17437412| ref|XP 0655	hypothetical protein XP 065554 [Hom(	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	gene_prediction1				54.1|	sapiens]	
B-lineage cells; mature granulocytes.	HG1000342 N0_160000_gene_prediction1	67	0.55	Transmembrane (TM).	gi|6753342|ref|NP_03397 6.1|	CD24a antigen; heat stable antigen [Mus musculus]	Disorders of hematopoiesis and immune system.
B-lineage cells; mature granulocytes.	HG1000342 N0_160000_gene_prediction2	67	0.55	TM.	gi|6753342|ref|NP_03397 6.1|	CD24a antigen; heat stable antigen [Mus musculus]	Disorders of hematopoiesis and immune system.
	HG1000650 N0_20000_gene_prediction1	77	0.23		gi|29250313|gb|EAA4180 9.1|	GLP_111_35594_437 26 [Giardia lamblia ATCC 50803]	
	HG1000191 N0_160000_gene_prediction2	50	0.92	Intracellular	gi|27479383|ref|XP_2080 98.1|	similar to RE52392p [Drosophila melanogaster] [Homo sapiens]	Diagnostic marker for therapeutic agents that upregulate this gene may result in beneficial effects in neurodegenerative diseases and delay symptom onset.
	HG1000449 N0_160000_gene_prediction3	94	0.85		gi|585328|sp|Q07654|TFF 3_HUMAN	Trefoil factor 3 precursor (Intestinal trefoil factor) (hP1.B)	
	HG1000181 N0_20000_g	1560	0.85		gi|11560152|ref|NP_0713	zinc finger protein 335; zinc-	

181

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ene_predicti on1				78.1|	finger/leucine-zipper co-transducer NIF1;	
	HG1001058 N0_160000_gene_predict ion1	135	0.17		gi|20826648| ref|XP_1609 59.1|	hypothetical protein XP_160959 [Mus musculus]	
	HG1000187 N0_160000_gene_predict ion2	54	0.22		gi|21225243| ref|NP_6310 22.1|	putative oxidoreductase. [Streptomyces coelicolor A3(2)]	
	HG1000191 N0_1000_ge ne_ prediction1	32	0.9		gi|27479383| ref|XP_2080 98.1|	similar to RE52392p [Drosophila melanogaster] [Homo sapiens]	
	HG1000319 N0_160000_gene_predict ion1	67				no_blastp_hit	
Highly expressed in fetal pancreas; broadly and equivalently expressed in bone marrow, spleen, brain, liver, pancreas, muscle, prostate, kidney,	HG1000137 N0_0_gene_ prediction1	64	0.84	TM.	gi|13124770| ref|NP_0768 69.1|	hypothetical protein IMAGE3455200 [Homo sapiens]	Diabetes; pancreatic cancer.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
lung.							
Expressed everywhere examined except thymus, spinal cord.	HG1000191 N0_5000_ge ne_ prediction1	75	0.38	ER	gi|27479383| ref|XP_2080 98.1|	similar to RE52392p [Drosophila melanogaster] [Homo sapiens]	Homologus to DNAJ, a chaperone.
	HG1001350 N0_5000_ge ne_ prediction1	540	0.09		gi|7512280|p ir||T03455	ALR protein - hur gi|2358287|gb|AAC51 735.1| ALR [Homo sapiens]	
	HG1000327 N0_160000_ gene_predict ion2	582	0.57	Secreted; lysosomal.	gi|27676136| ref|XP_2235 21.1|	similar to prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy):	Variant Gaucher disease and variant metachromatic leukodystrophy and other metabolic diseases; neuro.: metachromatic leuko- dystrophy variant (e.g., metachromatic leuko- dystrophy due to SAP-1 deficiency), neonatal hyperkinetic behavior (e.g., combined SAP deficiency), myoclonus (e.g., Combined SAP deficiency); resp.: respiratory insufficiency (e.g., combined SAP deficiency); GI: hepato- splenomegaly (e.g., combined SAP deficiency),

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000179N0_160000_gene_prediction1	221	0.57		gi|13436134|gb|AAH04884.1|AAH04884	Unknown (protein for MGC:11141) [Homo sapiens]	cher disease, atypical); lab: prosaposin deficiency, marked glucosylceramide accumulation in the spleen (e.g., Gaucher disease, atypical); inheritance: autosomal dominant (10q21-q22).
Brain; blood; dorsal root ganglion; thyroid; Unigene: human placenta; pooled glandular; pooled glioblastoma; normal epithelium; cochlea; pooled germ cell tumors; malignant melanoma (metastatic to lymph node); three pooled menin-	HG1000991N0_160000_gene_prediction1	237	0.66	Nuclear.	gi|6005864|ref|NP_009213.1|	ring finger protein 13; RING zinc finger protein [Homo sapiens]	Transcriptional activation; cancer; modulation of cell growth, cell differentiation, cell activation.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
giomas; two pooled squamous cell carcinomas; endometrioid ovarian metastasis; moderately-differentiated head and neck adenocarcinoma Neuromblastoma cot 50-normalized; pooled skin; blood vessels (aorta, basilar artery; placenta; placenta cot 50-normalized; colon; adipose; prostate; 7 pooled well-differentiated endometrial adenocarcinoma s; testis; bone marrow stroma; adenocarcinoma ; carcinoid;							

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
breast; follicular lymphoma; stomach; Kidney; pre-eclamptic placenta; lung; uterus; prostate; 2 pooled clear cell type tumors; poorly differentiated adenocarcinoma with signet ring cell features; liver; pooled human melanocyte, fetal heart, pregnant uterus; 2 pooled high-grade transitional cell tumors; ovary; normal breast; carcinoma cell line; corresponding non cancerous liver tissue; melanotic melanoma, high							

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
MDR; nervous tumor; adrenal gland; hyper-nephroma; nor-mal testis; un-differentiated large cell carcinoma; pit-uitary; bone marrow; whole embryo, mainly head; acute myelogenous leukemia; prim-itive neuro-ectoderm; skel-etal muscle; cervix; uterine tumor; normal Lung; trabecular meshwork; hypothalamus; glioblastoma with EGFR amplification; embryo; Colon_ins; normal plac-enta; duodenal							

187

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
adenocarcinoma cell line; mammary adeno-carcinoma cell line; adrenal cortex carcin-oma cell line; marrow; hyper-nephroma cell line; melano-cyte; squamous cell carcinoma; cartilage; liver; anaplastic oligo-dendroglioma with 1p/19q loss; hippocam-pus; multiple sclerosis les-ions; neuro-blastoma; skin; hypothalamus; retinoblastoma; pooled colon, kidney, stom-ach; amelanotic melanoma cell line; retina (foveal and							

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
macular); heart; high grade serous papillary carcinoma, 2 pooled tumors; leiomyosarcoma; native retinal pigment epithelium (RPE) sheets; human retina; stomach; lung; primary lung epithelial cells; metastatic chondrosarcoma; melanotic melanoma; grade II chondrosarcoma; fetal eyes, lens, eye anterior segment, optic nerve, RPE and choroid; sympathetic trunk; purified pancreatic islet; primary lung cystic fibrosis							

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
epithelial cells; chondrosarcoma; ascites; insu-linoma; aorta; para-thyroid tumor; leuk-ocyte; renal cell tumor; pineal gland; mixed pool of 40 RNAs; human lung epithelial cells; fibro-sarcoma; Ew-ing's sarcoma; alveolar rhabdo-myosarcoma; germinal center B cell; spleen. Brain.	HG1001038 N0_5000_ge ne_ prediction1	174	0.8	Integral membrane protein; ER.	gi|13279206| gb|AAH043 13.1|AAH04 313	Unknown (protein for IMAGE:3627962) [Homo sapiens]	Type IV of the carbohydrate deficient glycoprotein syndromes (CDGS) is characterized by microcephaly, severe epilepsy, minimal psycho-motor development and partial deficiency of sialic acids in serum glycol-

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1001376 N0_160000_ gene predict	652	0.81		gi|7512821|p ir||T00347	hypothetical protein DKFZp566G1246.1, version I - human	defect is a missense mutation in the gene encoding the mannosyltransferase

dolichyl-phosphate mannose on to the lipid-linked oligosaccharide (LLO) intermediate Man(5) GlcNAc(2)-PP-dolichol, resulting in the accumulation of the LLO intermediate and, due to its leaky nature, a residual formation of full-length LLOs; N-glycosylation is

oligosaccharides are transferred in addition to full-length oligosaccharides, and incomplete utilization of N-glycosylation sites; the mannosyltransferase is the str orthologue of the *Saccharomyces cerevisiae* (Korner et al., 1999). |

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
ion2						(fragment)	
	HG1001376 N0_20000_g ene_predicti on2	849	0.83		gi|22052701| ref|XP_0519 56.8|	similar to hypothetical protein DKFZp566G1246.1, version I - human (fragment) [Homo	
	HG1000409 N0_160000_ gene_predict ion1	260	0.71		gi|106322|pi r||B34087	hypothetical protein (L1H 3' region) - human	Hemophilia.
	HG1000884 N0_160000_ gene_predict ion1	289	0.61		gi|28478194| ref|XP_1333 89.3|	similar to hypothetical protein FLJ32949 [Mus sapiens] [Mus musculus]	
	HG1000575 N0_160000_ gene_predict ion1	79				no_blastp_hit	
	HG1000906 N0_10000_g ene_predicti on1	216	0.53		gi|4505843|r ef|NP_00361 9.1|	plakophilin 4 [Homo sapiens] gi|20139104|sp|Q9956 9|PKP4_HUMAN Plakophilin 4 (p0071)	
	HG1000485 N0_160000_ gene_predict ion1	312	1		gi|4927640|g b|AAD3321 5.1|	PPP1R5 [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000328 N0_160000_gene_prediction1	301	0.35		gi|7662198|ref|NP_05564 7.1|	TBC1 domain family, member 4; KIAA0603 gene product; TBC (Tre-2, BUB2, CDC16) domain-containing	Testing the toxicity of chemicals on human cells as an *in vitro* model of demyelinating disease; certain neuropathic disorders may correlate with inactivating mutations; levels and activity may serve as a useful marker; upregulation may provide protection against prolonged effects of chemical agents.
	HG1000231 N0_160000_gene_prediction1	89	0.92	Cytosolic.	gi|4679012|gb|AAD2699 4.1|	lysophospholipase isoform [Homo sapiens]	
Activated lymphocytes.	HG1001257 N0_10000_gene_prediction1	294	0.36	TM; possibly ER.	gi|18676478|dbj|BAB848 91.1|	FLJ00136 protein [Homo sapiens]	Epidermodysplasia verruciformis cancer, esp. skin carcinoma.
Ubiquitous; highly expressed in Bone Marrow, intermediate to high levels in	HG1000026 N0_5000_gene_prediction1	379	0.52	Integral membrane protein; mitochondrial inner membrane;	gi|12248755|dbj|BAB202 65.1|	mono ATP-binding cassette protein [Homo sapiens]	Anti-drug resistance; Tangier disease, a rare autosomal recessive disorder caused by mutations in ATP-binding

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
skeletal muscle, small intestine, thyroid, heart, brain, placenta, liver, pancreas, prostate, testis, ovary, leuko-cyte, stomach, spinal cord, lymph node, trachea and adrenal gland; low levels in lung, kidney, spleen, thymus and colon.				STM.			(ABC1); a typical clinical manifestation is peripheral neuropathy (Bodzioch et al., 1999).
	HG1000300 N0_160000_gene_predict ion1	188	0.81		gi|12804335| gb|AAH030 26.1|AAH03 026	Unknown (protein for IMAGE:2823490) [Homo sapiens]	
	HG1000109 N0_160000_gene_predict ion1	1837	0.23		gi|28570186| ref|NP_0606 18.2|	WINS1 protein [Homo sapiens] gi|21203151|dbj|BAB 93864.1| WINS1 protein with Drosophila Lines	Immunosupression and cancer therapeutics.
	HG1001110 N0_160000_gene_predict	1837	0.23		gi|28570186| ref|NP_0606 18.2|	WINS1 protein [Homo sapiens] gi|21203151|dbj|B,	

194

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ion1					93864.1| WINS1 protein with Drosophila Lines	
	HG1001376 N0_160000_ gene_predict ion3	874	0.9		gi|7512821|p ir||T00347	hypothetical protein DKFZp566G1246.1, version I - human (fragment)	
Ubiquitous; Overexpressed in mantle cell lymphoma, follicular lymphoma; highly expressed in bone marrow; expressed at intermediate to high levels in skeletal muscle, small intestine, thyroid, heart, brain, placenta, liver, pancreas, prostate, testis, ovary, leukocyte, stomach, spinal cord, lymph	HG1000026 N0_20000_g ene_predicti on1	452	0.44	Integral membrane protein; mitochondri al inner membrane.	gi|12248755| dbj|BAB202 65.1|	mono ATP-binding cassette protein [Homo sapiens]	Members of the usually transmembrane proteins, consisting minimally of 2 TM domains responsible for allocrite binding and transport; may be difficult to target when in inner mitochondrial membrane, so useful as diagnostic marker.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
node, trachea, adrenal gland expressed at low levels in lung, kidney, spleen, thymus, colon.							
High expression (EST) in primitive neuroectoderm, embryonal carcinoma, mucoepidermoid carcinoma from chronic myelogenous leukemia, glioblastoma, acute myelogenous leukemia, hypernephroma, melanotic melanoma.	HG1000276 N0_1000_gene_prediction1	131	0.54	Secreted (based on bioinformatics); STM, e.g., harvester; FP is 5' and 3' complete; 3' is longer than the reference sequence; identity 91% (71/78) (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=protein&list_uids=8923930&dopt=GenPept).	gi|8923930|ref|NP_06093 4.1|	uncharacterized hematopoietic stem/progenitor cells protein MDS029 [Homo sapiens]	Immune disorders, cancer, neuronal disorders.
Ubiquitous;	HG1000822	183	0.33	Nuclear.	gi|20883730|	similar to histone	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
higher levels in heart, pancreas testis; lower levels in kidney, brain.	NO_160000_gene_prediction2				ref|XP_1233 11.1|	deacetylase 1 [Mus musculus]	2003).
Testis.	HG1000173 NO_20000_gene_prediction1	271	0.84	Secreted'	gi|21751939| dbj|BAC040 76.1|	unnamed protein product [Homo sapiens]	
	HG1001044 NO_1000_gene_prediction1	131	0.26		gi|25020270| ref|XP_2078 43.1|	hypothetical protein XP_207843 [Mus musculus]	
	HG1000299 NO_1000_gene_prediction1	194	0.71		gi|4503729|r ef|NP_00200 5.1|	FK506-binding protein 4; FK506-binding protein 4 (59kD); T-cell FK506-binding protein, 59kD;	Immunosupression and cancer.
	HG1000659 NO_160000_gene_prediction1	257	0.28		gi|27498395| ref|XP_2106 62.1|	similar to hypothetical protein; sequence orphan; low similarity to glycoamylases and	
	HG1000659 NO_160000_gene_prediction2	257	0.28		gi|27498395| ref|XP_2106 62.1|	similar to hypothetical protein; sequence orphan; low similarity to glycoamylases and	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000013N0_160000_gene_prediction1	63				no_blastp_hit	
	HG1000173N0_160000_gene_prediction1	256	0.54	Secreted.	gi|21751939|dbj|BAC04076.1|	unnamed protein product [Homo sapiens]	
	HG1000330N0_160000_gene_prediction1	202	0.11		gi|23508369|ref|NP_701038.1|	hypothetical protein [Plasmodium falciparum 3D7]	
	HG1000178N0_10000_gene_prediction1	286	0.48		gi|22035592|ref|NP_057706.2|	mitochondrial ribosomal protein] isoform a [Homo sapiens]	
	HG1000178N0_10000_gene_prediction2	286	0.48		gi|22035592|ref|NP_057706.2|	mitochondrial ribosomal protein] isoform a [Homo sapiens]	
	HG1000640N0_160000_gene_prediction1	531	0.14		gi|15011990|gb|AAH10889.1|AAH10889	Similar to CGI-116 protein [Homo sapiens]	
	HG1001000N0_160000_gene_prediction1	280	0.37		gi|27483462|ref|XP_208106.1|	similar to hypothetical protein BC013073 [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1001418 N0_160000_gene_predict ion1	219	0.16		gi|20819462| ref|XP_1580 58.1|	hypothetical protein XP_158058 [Mus musculus]	
	HG1000153 N0_20000_g ene_predicti on1	1192	0.79		gi|20143967| ref|NP_6125 65.1|	kinesin-like 5 isoform 1; mitotic kinesin-like 1 [Homo sapiens]	
	HG1000255 N0_160000_gene_predict ion1	92	0.44		gi|3212110|e mb|CAA767 59.1|	prefoldin subunit 1 [Homo sapiens]	
	HG1000186 N0_160000_gene_predict ion1	63	0.34		gi|10092615| ref|NP_0611 08.2|	ethanolamine kinase [Homo sapiens] gi|14194724|sp|Q9HB U6|EKI1_HUMAN Ethanolamine kinase (EKI)	
High expression in brain, spinal cord (EST).	HG1000259 N0_160000_gene_predict ion1	73	0.5	Stathmin-like-protein, e.g., T08783 Stathmin 4 (189 aa).	gi|7512937|p ir||T08783	hypothetical protein DKFZp586O0120.1 - human (fragment)	Neuronal disorders, neurodegenerative disorders such as Alzheimer's disease (Mori, 1997).
	HG1000084 N0_10000_g ene_predicti on1	244	0.17		gi|6690017|g b|AAF23950 .1|	NTR [Herpesvirus papio]	
	HG1000217	299	0.09		gi|18577116|	hypothetical prote	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_160000_gene_prediction1				ref|XP_0845 21.1|	XP_084521 [Homo sapiens]	
	HG1000217 N0_160000_gene_prediction2	299	0.09		gi|18577116| ref|XP_0845 21.1|	hypothetical protein XP_084521 [Homo sapiens]	
	HG1000329 N0_160000_gene_prediction1	238	0.2		gi|22962211| gb|ZP_0000 9817.1|	hypothetical protein [Rhodopseudomonas palustris]	
	HG1000227 N0_160000_gene_prediction1	132	0.9	Mitochondrial inner membrane.	gi|4506863|r ef|NP_00299 2.1|	succinate dehydrogenase complex, subunit C precursor; Succinate dehydrogenase complex,	
	HG1000269 N0_10000_gene_prediction1	132	0.43		gi|7706341|r ef|NP_05714 5.1|	yippee protein [Homo sapiens] gi|20901030|ref|XP_1 28760.1| RIKEN cDNA 2310076K21 [Mus musculus]	
	HG1000990 N0_160000_gene_prediction1	354	0.64		gi|8924262|r ef|NP_06111 3.1|	triggering receptor expressed on myeloid cells 1 [Homo sapiens]	
	HG1000998	89	0.15		gi|26541476|	alpha-TIF [canine	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_160000_gene_prediction1				gb|AAN854 77.1|AF3886 72_2	herpesvirus 1]	
	HG1001225 N0_160000_gene_prediction1	123	0.09		gi|2498888|s p|Q28892|S5 A2_MACF A	3-oxo-5-alpha-ster 4-dehydrogenase 2 (Steroid 5-alpha-reductase 2) (SR type	
Ubiquitous.	HG1001269 N0_5000_ge ne_ prediction1	132	0.53	Secreted.	gi|5231228|r ef|NP_00372 1.2|	ribonuclease 6 precursor [Homo sapiens] gi|20139363|sp|O0058 4|RNP6_HUMAN Ribonuclease 6 precursor	
	HG1001269 N0_160000_gene_predict ion1	167	0.41		gi|5231228|r ef|NP_00372 1.2|	ribonuclease 6 precursor [Homo sapiens] gi|20139363|sp|O0058 4|RNP6_HUMAN Ribonuclease 6 precursor	
	HG1001103 N0_160000_gene_predict ion1	453	0.74		gi|4507677|r ef|NP_00329 0.1|	tumor rejection antigen (gp96) 1; Tumor rejection antigen-1 (gp96) [Homo sapiens]	
	HG1000143 N0_1000_ge ne	119	0.74		gi|21594169| gb|AAH317 46.1|	Similar to ribosomal protein S9 [Mus musculus]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
prediction1							
	HG1000396 N0_160000_ gene_predict ion1	369	0.83		gi|1196433|g b|AAA8803 8.1|	unknown protein	
	HG1000066 N0_160000_ gene_predict ion1	80	0.46		gi|28376635| ref|NP_7838 65.1|	Rab37-like [Homo sapiens] gi|26252126|gb|AAH 40547.1| Similar to RAB37, member of RAS	Cancer, esp. melanoma, neuroblastoma.
	HG1000078 N0_1000_ge ne_ prediction1	85	0.72	Nuclear.	gi|27663408| ref|XP_2165 94.1|	similar to cell division cycle 2-like 1, isoform 1; Cell division cycle 2-like 1; PITSLRE	
	HG1000117 N0_160000_ gene_predict ion1	172	0.83		gi|7512733|p ir||T08691	hypothetical protein DKFZp564F052.1 - human (fragment)	
	HG1000157 N0_160000_ gene_predict ion1	151	0.21		gi|24647887| ref|NP_7323 10.1|	CG31196-PB [Drosophila melanogaster] gi|23171616|gb|AAN 13764.1| CG31196-PB [Drosophila melanogaster]	
	HG1000194	98	0.34		gi|7662639|r	PTD011 protein	

202

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_160000_gene_prediction1				ef|NP_05477 0.1|	[Homo sapiens] gi|9910842|sp|Q9Y6G1|PD11_HUMAN Protein PTD011 (Protein TS58)	
	HG1000228 N0_40000_gene_prediction1	111	0.15		gi|15607313| ref|NP_2146 86.1|	hypothetical protein Rv0172 [Mycobacterium tuberculosis H37Rv]	
	HG1000228 N0_20000_gene_prediction1	92	0.18		gi|15607313| ref|NP_2146 86.1|	hypothetical protein Rv0172 [Mycobacterium tuberculosis H37Rv]	
	HG1000228 N0_160000_gene_prediction1	128	0.13		gi|15607313| ref|NP_2146 86.1|	hypothetical protein Rv0172 [Mycobacterium tuberculosis H37Rv]	
	HG1000409 N0_10000_gene_prediction1	89	0.21		gi|116138|sp| P11651|CFI B_PETHY	Chalcone--flavonone isomerase B (Chalcone isomerase B)	
	HG1000611 N0_160000_gene_prediction1	159	0.16		gi|21299803| gb|EAA1194 8.1|	agCP6635 [Anopheles gambiae str. PEST]	
Developmentally-regulated membrane-	HG1000015 N0_0_gene_prediction1	1194	0.69	STM	gi|13591932| ref|NP_1122 84.1|	membrane-spanning proteoglycan NG2 [Rattus norvegicus]	Cancer, tumor targeting.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
spanning chondroitin sulfate proteoglycans;expressed primarily by glial, muscle, and cartilage progenitor cells; upon maturation, these cell types down-regulate NG2 expression; in adult animals, the expression of NG2 is restricted to tumor cells and angiogenic tumor vasculature, making this proteoglycan a potential target for directing therapeutic agents to relevant sites of action.							
M1 type is in	HG1000088	317	0.87	Intracellular	gi|2117873|p	pyruvate kinase (E	Metabolic disorders.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
muscle, liver, red cell, fetal tissue; L type is major liver isozyme; R type is in red cells; M1 is main form in muscle, heart, brain; M2 is early fetal tissues.	N0_5000_ge ne_ prediction1			(cytoplasm).	ir||S64635	2.7.1.40), muscle splice form M1 - human	
	HG1000143 N0_10000_g ene_predicti on1	408	0.67		gi|4502601|r ef|NP_00122 7.1|	carbonyl reductase 3; carbonyl reductase (NADPH) 3 [Homo sapiens]	
GeneCard (based on KIAA1143): high abundacy in bone marrow; similar levels in brain, liver, prostate, kidney, heart, pancreas, lung; based on KIAA1143: eye, pheochromocyt oma, high grade prostatic.	HG1000167 N0_5000_ge ne_ prediction1	158	0.5		gi|16877034| gb|AAH167 90.1|AAH16 790	Unknown (protein for MGC:24015) [Homo sapiens]	Diagnostic marker for preinvasive stage of adenocarcinoma, e.g. high-grade prostatic intra-epithelial neoplasia (PIN) (has a high predictive marker value for adenocar-cinoma, identification warrants repeat biopsy for concurrent or subsequent invasive carcinoma).

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
intraepithelial neoplasia.							
	HG1000243 N0_5000_ge ne_ prediction1	127	0.7		gi|18555712| ref|XP_0961 98.1|	hypothetical protein XP_096198 [Homo sapiens]	
	HG1000825 N0_160000_ gene_predict ion1	105	0.29		gi|23490837| gb|EAA2251 7.1|	putative methionyl-tRNA synthetase [Plasmodium yoelii yoelii]	
	HG1001019 N0_1000_ge ne_ prediction1	63	0.49		gi|27479097| ref|XP_2099 31.1|	similar to hypothet protein, 1-107; hypothetical protein, clone 1-107 [Mus musculus]	
	HG1000044 N0_160000_ gene_predict ion1	310	0.91	Intracellular.	gi|13431718| sp|Q9ULV0| MY5B_HU MAN	Myosin Vb (Myosin 5B) gi|6329708|dbj|BAA8 6433.1|KIAA1119 protein [Homo sapiens]	Heart failure, microtubule-disrupting disorders.
	HG1000100 N0_10000_g ene_predicti on1	114	0.86	Intracellular (cytoplasm).	gi|4506127|r ef|NP_00275 5.1|	phosphoribosyl pyrophosphate synthetase 1 [Homo sapiens]	Familial disorder characterized by excessive purine production, gout, and uric acid urolithiasis, (Roessler et al., 1993).
	HG1000149	89	0.21		gi|21244089||	conserved	

206

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	NO_160000_gene_predict ion1				ref|NP_6436 71.1|	hypothetical protein [Xanthomonas axonopodis pv. citri str. 306]	Diagnostic monitoring of protein levels, mRNA, or mutations in the gene that encodes for BAF may pre agents that affect BAF levels may alter the course of the disorder.
	HG1000183 NO_1000_ge ne_ prediction1	134	0.75	Cytoplasmic	gi|18148873| dbj|BAB835 17.1|	hUST3 [Homo sapiens]	
	HG1000183 NO_160000_ gene_predict ion2	134	0.75		gi|18148873| dbj|BAB835 17.1|	hUST3 [Homo sapiens]	
	HG1000213 NO_5000_ge ne_ prediction1	73	0.64	Nuclear.	gi|4502389|r ef|NP_00385 1.1|	barrier to autointegration fac Breakpoint cluster region protein, utel leiomyoma, 1	
	HG1000294 NO_5000_ge ne_ prediction1	190	0.56		gi|11386175| ref|NP_0687 78.1|	protein phosphatase 1, regulatory (inhibitor) subunit 11 isoform 1; hemochromatosis	
	HG1000430 NO_160000_ gene_predict ion1	115	0.13		gi|14754198| ref|XP_0454 50.1|	similar to hypothetical protein (L1H 3 region) - human [Homo	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000078 N0_5000_gene_prediction1	109	0.55		gi|27663408| ref|XP_2165 94.1|	similar to cell division cycle 2-like 1, isoform 1; Cell division cycle 2-like 1; PITSLRE sapiens]	
	HG1000139 N0_5000_gene_prediction1	142	0.77		gi|13375725| ref|NP_0788 34.1|	chromosome 14 o reading frame 138 [Homo sapiens]	
	HG1000143 N0_160000_gene_predict ion1	447	0.61		gi|4502601|r ef|NP_00122 7.1|	carbonyl reductase 3; carbonyl reductase (NADPH) 3 [Homo sapiens]	
The ribosomal protein L13e is widely expressed in vertebrates, Drosophila melanogaster, plants, yeast, others.	HG1000162 N0_160000_gene_predict ion1	193	0.71	Cytoplasmic	gi|15431295| ref|NP_1502 54.1|	ribosomal protein L13; 60S ribosomal protein L13; breast basic conserved protein 1 [Homo	
High abundance.	HG1000168 N0_160000_gene_predict ion1	90	0.77		gi|1710488|s p|P50914|RL 14_HUMA N	60S ribosomal protein L14 (CAG-ISL 7)	
	HG1000187	338	0.98		gi|2072963|g	p40 [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_160000_gene_prediction1				b|AAC51270.1|		
	HG1000247 N0_160000_gene_prediction1	302	0.85		gi|4757824|ref|NP_004646.1|	axin 2; axil [Homo sapiens] gi|12643949|sp|Q9Y2T1|AXN2_HUMAN Axin 2 (Axis inhibition protein 2)	
	HG1000273 N0_160000_gene_prediction2	125	0.2		gi|10438240|dbj|BAB15204.1|	unnamed protein product [Homo sapiens]	
	HG1000539 N0_160000_gene_prediction1	205	0.14		gi|4007653|emb|CAA05243.1|	exo-1,3-beta-glucanase [Pichia anomala]	
	HG1000539 N0_160000_gene_prediction2	205	0.14		gi|4007653|emb|CAA05243.1|	exo-1,3-beta-glucanase [Pichia anomala]	
	HG1000560 N0_160000_gene_prediction1	176	0.16		gi|25032591|ref|XP_111679.2|	similar to 573K1.8 (mm17M1-2 (novel 7 transmembrane receptor (rhodopsin family) (olfactory family)	
	HG1000740 N0_160000	370	0.11		gi|28921429|gb|EAA3073	predicted protein [Neurospora crassa]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	gene_prediction1				5.1|		
	HG1000020 N0_5000_gene_prediction1	214	0.22		gi|7263959|emb|CAB81642.1|bA395L14.5	(novel phosphoglucomutase like protein) [Homo sapiens]	
	HG1000084 N0_5000_gene_prediction1	192	0.18		gi|4520376|dbj|BAA7591 3.1|	uridylyl transferase [Pseudomonas aeruginosa]	
	HG1000135 N0_5000_gene_prediction1	114	0.93		gi|13569911| ref|NP_1122 02.1|	hypothetical protein MGC4276 similar to CG8198 [Homo sapiens]	
	HG1000169 N0_20000_gene_prediction1	314	0.78	Mitochondrial.	gi|119522|sp| P10658|SER C_RABIT	phosphoserine aminotransferase (psat) (endometrial progesterone-induced protein) (epip)	
	HG1000169 N0_160000_gene_prediction1	314	0.78		gi|119522|sp| P10658|SER C_RABIT	phosphoserine aminotransferase (psat) (endometrial progesterone-induced protein) (epip)	
	HG1000189 N0_160000_gene_prediction1	282	0.31		gi|27684635| ref|XP_2261 10.1|	similar to BG:DS01759.1 gene product [Drosophila melanogaster] [Rat	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000189 N0_160000_gene_prediction2	282	0.31		gi|27684635|ref|XP_2261 10.1|	similar to BG:DS01759.1 gene product [Drosophila melanogaster] [Rattus norvegicus]	norvegicus]
	HG1000246 N0_5000_gene_prediction1	80	0.43		gi|5630076|gb|AAD4582 1.1|AC0060 17 1|	N-acetylgalactos-aminyltransferase; similar to Q10473	
Widely expressed in various tissues.	HG1000248 N0_0_gene_prediction1	131	0.81		gi|5802966|ref|NP_00686 1.1|	destrin (actin depolymerizing factor); destrin [H sapiens]	
	HG1000288 N0_10000_gene_prediction1	208	0.45		gi|1263081|gb|AAC5201 0.1|	mariner transposase [Homo sapiens]	
	HG1000443 N0_40000_gene_prediction1	194	0.2		gi|28972658|dbj|BAC657 45.1|	mKIAA1197 protein [Mus musculus]	
	HG1000590 N0_1000_gene_prediction1	140	0.35		gi|26378096|dbj|BAB285 95.2|	unnamed protein product [Mus musculus]	
	HG1000626 N0_160000	92	0.2		gi|14030530|gb|AAK529	pol polyprotein [Human	

211

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	gene_predict ion1				38.1|	immunodeficiency virus type 1]	
	HG1000871 N0_160000_gene_predict ion1	713	0.65		gi|3915750|s p|P37023|KI R3_HUMA N	Serine/threonine-protein kinase receptor R3 precursor (SKR3) (Activin receptor-like	Extracellular domain, including variable part thereof, as antibody target; palmitoylation of intracellular domain; inhibitors useful for lipid modification; neutralizing antibodies useful for modulting virus entry.
Family members expressed in every cell type examined; specificity for individual tetraspanins displayed; FP clone displays similiarities to a platelet tetraspanin CD 151.	HG1000959 N0_10000_g ene_predicti on1	419	0.3	Plasma membrane; endosomal and lysosomal membranes.	gi|27713370| ref|XP_2323 71.1|	similar to RIKEN cDNA 1110014F12 [Mus musculus] [Rattus norvegicus]	
	HG1000961 N0_160000_gene_predict ion3	460	1		gi|3115348|g b|AAC1586 6.1|	mitochondrial processing peptidase beta subunit precursor; similar to Q03346 (PID:g1171010)	
CD34+ stem cells; dendritic	HG1000974 N0_5000_ge	85	0.89		gi|6841264|g b|AAF28985	HSPC307 [Homo sapiens]	Cancer: modulation of cell

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
cells.	ne_prediction1				.1|AF161425_1		eration; cell differentiation; modulation of chronic inflammation: therapeutic for autoimmunity, including RA, SLE, MS, insulin-dependent diabetes
	HG1001045 N0_160000_gene_predict ion1	271	0.83	Nuclear.	gi|5070621|gb|AAD3921 4.1|AF14885 6_1	unknown [Homo sapiens]	cancer therapy: dendritic cell sarcoma, cancer immunotherapy, cancer vaccines, growth and/or differentiation of immature or mature blood cells; regeneration of cells of HSC lineage; bone marrow transplantation; *ex vivo* or *in vivo* of HSC lineage. Genetic diseases.
Brain, heart, prostate, lung; high levels in adult and fetal brain, kidney, lung; low levels in adult placen-	HG1001110 N0_0_gene_prediction1	110	0.95		gi|10835065|ref|NP_0027 51.1|	protein kinase, Y-linked [Homo sapiens]	PRKX, a phylogenetically and functionally distinct cAMP-dependent protein kinase, (Page et al, 1999; van der Spoel et al., 2002); variant in cancer.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
ta, heart, liver, skeletal muscle, pancreas, fetal liver.							
	HG1001223 N0_1000_ge ne_ prediction1	65	1		gi|21040253| ref|NP_6319 06.1|	sarcoglycan zeta; zeta-sarcoglycan [Homo sapiens]	
	HG1001281 N0_160000_ gene_predict ion1	361	0.54		gi|27481712| ref|XP_0479 61.5|	similar to HCH [Mus musculus] [Homo sapiens]	
	HG1001317 N0_5000_ge ne_ prediction1	144	0.66		gi|26327365| dbj|BAC274 26.1|	unnamed protein product [Mus musculus]	
	HG1001017 N0_10000_g ene_predicti on1	92	0.67		gi|25021133| ref|XP_2077 37.1|	hypothetical protein XP_207737 [Mus musculus]	
	HG1001017 N0_1000_ge ne_ prediction1	84	0.71		gi|25021133| ref|XP_2077 37.1|	hypothetical protein XP_207737 [Mus musculus]	
	HG1000014 N0_160000_ gene_predict ion2	600	0.36		gi|4502281|r ef|NP_00167 0.1|	ATPase, Na+/K+ transporting, beta 3 polypeptide [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000043 N0_160000_gene_prediction3	519	0.69		gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor D1B; D1beta dopamine receptor [Homo sapiens]	
	HG1000052 N0_160000_gene_prediction1	1129	0.9		gi|5052951|gb|AAD38785.1|AF149422_2	unknown [Homo sapiens]	
	HG1000084 N0_5000_gene_prediction2	161	0.29		gi|27478305|ref|XP_209869.1|	hypothetical protein XP_209869 [Homo sapiens]	
Ubiquitous; high expression in invasive ductal carcin-oma and liver.	HG1000093 N0_1000_gene_prediction1	54	0.9	Nuclear--protein kinase.	gi|1399462|gb|AAB03268.1|	serine/threonine-protein kinase PRP4h	
	HG1000105 N0_160000_gene_prediction1	282	0.45		gi|4757930|ref|NP_004692.1|	cyclin B2 [Homo sapiens] gi|5921731|sp|O95067|CGB2_HUMAN G2/mitotic-specific cyclin B2	
	HG1000157 N0_1000_gene_prediction1	118	0.52		gi|284658|pir||S23303	protein kinase C inhibitor KCIP-1 isoform epsilon - sheep	Myotonic dystrophy (Steinert disease), a protein splicing disorder.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
Kidney, lung, brain, pancreas, skeletal muscle.	HG1000210 N0_40000_gene_prediction1	591	0.43	PDZ domain proteins are frequently associated with the plasma membrane.	gi|21450785| ref|NP_0042 23.2|	suppressor of cytokine signaling 4; STAT induced STAT inhibitor-4; cytokine-inducible SH2	Cytokine-induced STAT inhibitor (CIS), a suppressor of cytokine signaling (SOCS) in STAT-induced STAT inhibitor (SSI), protein family; members are known to be cytokine-inducible negative regulators of cytokine signaling; gene expression can be induced by GM- and EPO in hematopoietic cells; highly expressed in factor-independent chronic myelogenous leukemia (CML) and erythro-leukemia (HEL) cell lines.
	HG1000242 N0_5000_ge ne_prediction1	63	0.74		gi|14091768| ref|NP_1144 68.1|	DnaJ (Hsp40) homolog, subfamily A, member 2 [Rattus norvegicus]	
	HG1000243 N0_5000_ge ne_prediction2	70	0.94		gi|5031749|r ef|NP_00550 8.1|	high-mobility group nucleosomal binding domain 2; nonhistone chromosomal protein	
	HG1000256 N0_160000_gene_predict ion1	49	0.2		gi|24380252| ref|NP_7222 07.1|	conserved hypothetical protein [Streptococcus mutans UA159]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000279 N0_0_gene_ prediction1	162	0.88	Intracellular.	gi|14251209| ref|NP_0012 79.2|	chloride intracellular channel 1; p64CLCP [Homo sapiens]	Small molecule target for depression, obesity, schizophrenia, and osteoporosis.
	HG1000280 N0_5000_ge ne_ prediction1	495	0.72		gi|28514322| ref|XP_1097 34.2|	RIKEN cDNA 4732407F15 gene [Mus musculus]	
	HG1000280 N0_5000_ge ne_ prediction2	495	0.72		gi|28514322| ref|XP_1097 34.2|	RIKEN cDNA 4732407F15 gene [Mus musculus]	
	HG1000282 N0_160000_ gene_predict ion1	118	0.83		gi|9910382|r ef|NP_06462 8.1|	mitochondrial import receptor Tom22 [Homo sapiens]	
	HG1000292 N0_160000_ gene_predict ion1	434	0.14		gi|13639392| ref|XP_0176 61.1|	similar to ribosomal protein S26 [Homo sapiens]	
	HG1000313 N0_160000_ gene_predict ion1	157	0.85	Nuclear and/or cytoplasmic.	gi|20833586| ref|XP_1248 08.1|	similar to protein tyrosine phosphatase 4a1 [Rattus norvegicus] [Mus musculus]	Cancer, disorders of immune function.
	HG1000330 N0_20000_g ene predicti	64	0.26		gi|27664552| ref|XP_2167 09.1|	similar to hypothetical protein MGC30562 [Mus	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	on1					musculus] [Rattus norvegicus]	
	HG1000482 N0_160000_gene_predict ion1	384	0.11		gi|17484447| ref|XP_0661 02.1|	ribosomal protein L7a-like 2 [Homo sapiens]	
	HG1000486 N0_20000_g ene_predicti on1	87				no_blastp_hit	
	HG1000518 N0_160000_ gene_predict ion1	100	0.55		gi|21302998| gb|EAA1514 3.1|	agCP4709 [Anopheles gambiae str. PEST]	
	HG1000556 N0_160000_ gene_predict ion1	319	0.89		gi|106322|pi r||B34087	hypothetical protein (L1H 3' region) - human	
	HG1000588 N0_160000_ gene_predict ion1	147	0.14		gi|20178274| sp|O95782|A 2A1_HUM AN	Adapter-related protein complex 2 alpha 1 subunit (Alpha-adaptin A) (Adaptor protein	
	HG1000600 N0_160000_ gene_predict ion1	354	0.06		gi|22974415| gb|ZP_0002 0671.1|	hypothetical protein [Chloroflexus aurantiacus]	
	HG1000648	123	0.21		gi|22202761||	P0039H02.4 [Ory:	

218

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	NO_160000_gene_prediction1				dbj|BAC074 17.1|	sativa (japonica cultivar-group)]	
	HG1000696 NO_160000_gene_prediction1	422	0.16		gi|21758619| dbj|BAC053 39.1|	unnamed protein product [Homo sapiens]	
Brain (medulla oblongata).	HG1000788 NO_160000_gene_prediction1	828	0.87	Plasma membrane.	gi|27662804| ref|XP_2219 54.1|	similar to slit homolog 1; Slit1 [Rattus norvegicus]	Brain, anaplastic oligodendroglioma, neuronal regeneration, cancer.
	HG1000874 NO_160000_gene_prediction1	505	0.35		gi|13129102| ref|NP_0770 02.1|	hypothetical protein MGC955 [Homo sapiens]	
	HG1000902 NO_20000_gene_prediction1	312	0.53		gi|27684795| ref|XP_2372 60.1|	similar to chaperonin subunit 6a (zeta); chaperonin contain TCP-1 [Mus musculus]	
	HG1000902 NO_160000_gene_prediction2	312	0.54		gi|27684795| ref|XP_2372 60.1|	similar to chaperonin subunit 6a (zeta); chaperonin contain TCP-1 [Mus musculus]	
	HG1000902 NO_1000_gene	176	0.89		gi|14517632| dbj|BAB610. 32.1|	acute morphine dependence related protein 2 [Homo	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	prediction1					sapiens]	
	HG1000966 N0_1000_ge ne_ prediction1	127	0.7		gi|23270917| gb|AAH168 49.1|	Similar to hypothetical protein MGC25511 [Homo sapiens]	
	HG1000966 N0_5000_ge ne_ prediction1	127	0.7		gi|23270917| gb|AAH168 49.1|	Similar to hypothetical protein MGC25511 [Homo sapiens]	
	HG1000994 N0_160000_ gene_predict ion1	197	0.44		gi|27498059| ref|XP_0681 66.3|	similar to olfactory receptor MOR145-2 [Mus musculus] [Homo sapiens]	
Neurons in brain limbic regions; 10-fold higher affinity for dopamine than the D1 subtype; related pseudogenes reside on chro- mosomes 1 & 2.	HG1001014 N0_160000_ gene_predict ion3	519	0.69	Integral membrane protein.	gi|4503391|r ef|NP_00078 9.1|	dopamine receptor D5; dopamine receptor D1B; D1beta dopamine receptor [Homo sapiens]	Defects in drd5 cause blepharospasm, a primary focal dystonia affecting the orbicularis oculi muscles; symptoms include eye irritation and frequent blinking, progressing to involuntary spasms of eyelid closure; severe cases can lead to functional blindness.
	HG1001041 N0_5000_ge ne_ prediction1	122	0.45	Nuclear; Golgi.	gi|2136035|p ir||I38138	protein-serine kinase (EC 2.7.1.-) PSK-H1 - human (fragment)	

220

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1001337 N0_160000_gene_predict ion1	456	0.33		gi|7019521|r ef|NP_03748 4.1|	squamous cell carcinoma antigen recognized by T cells 2 [Homo sapiens]	
	HG1000151 N0_160000_gene_predict ion1	1112	0.51		gi|27665740| ref|XP_2343 03.1|	similar to RIKEN cDNA 1810048J11 [Mus musculus] [Rattus norvegicus]	
	HG1000330 N0_160000_gene_predict ion3	316	0.15		gi|25019980| ref|XP_2077 11.1|	hypothetical protein XP_207711 [Mus musculus]	
	HG1000957 N0_20000_g ene_predicti on1	728	0.54		gi|106322|pi r||B34087	hypothetical protein (L1H 3' region) - human	Biotechnology; gene cloning; directed evolution; genetic engineering; generation of recombinant DNA; generation of transgenic, knock-out or chimeric organisms.
	HG1000960 N0_0_gene_ prediction1	636	0.44		gi|20908689| ref|XP_1274 49.1|	RIKEN cDNA 4632401C08 [Mus musculus] gi|12852467|dbj|BAB 29422.1| unnamed protein product [Mus musculus]	
Neuronal.	HG1000960 N0_0_gene_ prediction2	603	0.8	Cell surface.	gi|20908689| ref|XP_1274 49.1|	RIKEN cDNA 4632401C08 [Mus musculus]	Parkinson's disease; Tourette's syndrome;

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1001280 N0_20000_g ene_predicti on1	1767	0.37			gi|12852467|dbj|BAB 29422.1| unnamed protein product [Mus musculus]	Hyperactivity Disorder; substance abuse treatment.
	HG1000003 N0_10000_g ene_predicti on1	142	0.24		gi|27477706| ref|XP_0461 26.4|	similar to Meningioma-expressed antigen 6/11 (MEA6) (MEA11) [Homo sapiens]	
	HG1000041 N0_160000_ gene_predict ion1	572	0.43		gi|27708380| ref|XP_2288 58.1|	similar to KIAA1280 protein [Homo sapiens] [Rattus norvegicus]	
	HG1000043 N0_160000_ gene_predict ion2	487	0.74		gi|27659604| ref|XP_2265 63.1|	similar to tripartite motif protein 9, isoform 1; homolog of rat RING finger Spring [Homo sapiens]	
	HG1000044 N0_5000_ge ne_ prediction1	222	0.49		gi|4503391|r ef|NP_00078 9.1|	dopamine receptor D5; dopamine receptor D1B; D1beta dopamine receptor [Homo sapiens]	
					gi|13431718| sp|Q9ULV0| MY5B_HU MAN	Myosin Vb (Myosin 5B) gi|6329708|dbj|BAA8 1119 protein [Homo	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000051 N0_160000_gene_predict ion1	844	0.54		gi|5453700|r ef|NP_00613 8.1|	interferon regulatory factor 6 [Homo sapiens] sapiens]	
expressed in many cell types in a complex with monomeric actin in a 1:1 ratio.	HG1000057 N0_160000_gene_predict ion1	99	0.71	Intracellular.	gi|130976|sp| P02584|PRO 1_BOVIN	Profilin I gi|625306|pir||FABO profilin - bovine	
Restricted to cells of neurologic origin.	HG1000060 N0_160000_gene_predict ion1	367	0.64	Intracellular.	gi|21293316| gb|EAA0546 1.1|	agCP10585 [Anopheles gambiae str. PEST]	Disorder of neuronal systems.
	HG1000061 N0_10000_g ene_predicti on1	157	0.52	Nuclear.	gi|6912244|r ef|NP_03622 9.1|	ADP-ribosylation factor-like 5 [Homo sapiens]	
Skeletal muscle.	HG1000079 N0_160000_gene_predict ion1	182	0.78	Mitochondri al (Noma et al., 2001).	gi|19923437| ref|NP_0573 66.2|	adenylate kinase 3 alpha like [Homo sapiens]	
	HG1000098 N0_160000_gene_predict ion1	945	0.55	STM (type I); secreted.	gi|20514776| ref|NP_6206 11.1|	guanylyl cyclase v kinase-like domain, soluble [Rattus norvegicus]	Myocardial infarction (Frithz; 1976); marker for prostate cancer (Hall et al., 1985); metabolic disease.
	HG1000105	214	0.6		gi|4757930|r	cyclin B2 [Homo	

223

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_5000_ge ne_ prediction1				ef|NP_00469 2.1|	sapiens] gi|5921731|sp|O95067 |CGB2_HUMAN G2/mitotic-specific cyclin B2	
	HG1000121 N0_160000_ gene_predict ion1	120	0.26		gi|7706497|r ef|NP_05739 2.1|	UMP-CMP kinase [Homo sapiens] gi|6578133|gb|AAF17 709.1|AF070416_1 UMP-CMP kinase [Homo sapiens]	
	HG1000131 N0_160000_ gene_predict ion1	294	0.2		gi|9507123|r ef|NP_06224 4.1|	factor-responsive smooth muscle protein [Rattus norvegicus]	
	HG1000134 N0_160000_ gene_predict ion1	512	0.27		gi|27479840| ref|XP_2082 77.1|	similar to heterogeneous nuc ribonucleoprotein C (C1/C2) [Homo sapiens]	
	HG1000134 N0_160000_ gene_predict ion2	512	0.27		gi|27479840| ref|XP_2082 77.1|	similar to heterogeneous nuc ribonucleoprotein C (C1/C2) [Homo sapiens]	
	HG1000136 N0_160000_ gene_predict ion1	101	0.85		gi|3342000|g b|AAC3991 2.1|	H beta 58 homolog [Homo sapiens] gi|9622852|gb|AAF89 954.1|AF175266_1	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
						vacuolar sorting protein 26	
	HG1000147 NO_160000_gene_prediction1	236	0.77		gi|13904870| ref|NP_0010 00.2|	ribosomal protein 40S ribosomal prot S5 [Homo sapiens]	
	HG1000166 NO_160000_gene_prediction1	929	0.35		gi|169337|s p|P31040|D HSA_HUM AN	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	
	HG1000172 NO_1000_gene_prediction1	96	0.9	Mitochondrial.	gi|6681095|r ef|NP_03183 4.1|	cytochrome c, somatic [Mus musculus] gi|6978725|ref|NP_03 6971.1| cytochrom somatic; Cytochrome C,	
	HG1000172 NO_1000_gene_prediction2	96	0.9		gi|6681095|r ef|NP_03183 4.1|	cytochrome c, somatic [Mus musculus] gi|6978725|ref|NP_03 6971.1| cytochrom somatic; Cytochrome C,	
	HG1000175 NO_5000_gene_prediction1	130	0.2		gi|23102869| gb|ZP_0008 9366.1|	hypothetical protein [Azotobacter vinelandii]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000175N0_10000_gene_prediction1	191	0.21		gi|14600870|ref|NP_147394.1|	hypothetical protein [Aeropyrum pernix]	
	HG1000175N0_160000_gene_prediction1	272	0.15		gi|14600870|ref|NP_147394.1|	hypothetical protein [Aeropyrum pernix]	
	HG1000175N0_1000_gene_prediction1	130	0.2		gi|23102869|gb|ZP_00089366.1|	hypothetical protein [Azotobacter vinelandii]	
	HG1000192N0_160000_gene_prediction1	270	0.14		gi|20140802|sp|Q9GZL7|WD12_HUMAN	WD-repeat protein (YTM1 homolog)	
	HG1000193N0_160000_gene_prediction2	249	0.53		gi|27477958|ref|XP_209657.1|	similar to RH38554p [Drosophila melanogaster] [Homo sapiens]	
	HG1000195N0_160000_gene_prediction1	260	0.21		gi|17390530|gb|AAH18231.1|	Unknown (protein for MGC:19236) [Mus musculus]	
	HG1000197N0_160000_gene_prediction1	392	0.07		gi|4106828|gb|AAD03033.1|	unknown [Rattus norvegicus]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000202 N0_20000_g ene_predicti on1	489	0.25		gi|28481020| ref|XP_1299 72.3|	similar to hypothetical protein [Homo sapiens] [Mus musculus]	
	HG1000210 N0_20000_g ene_predicti on1	412	0.62	Cytoplasmic	gi|21450785| ref|NP_0042 23.2|	suppressor of cytokine signaling 4; STAT induced STAT inhibitor-4; cytokine-inducible SH2	Inflammation, viral replication, heart disease.
cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression: human mamm-ary luminal epithelial cells; differentially regulated genes include those involved in cell-matrix inter-actions, inclu-ding proline 4-hydroxylase (P4HA2), galectin 1,	HG1000218 N0_1000_ge ne_ prediction1	76	0.94	Cytoplasmic	gi|6681015|r ef|NP_03178 9.1|	cysteine rich intestinal protein [Mus musculus]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
(LGALS1), galectin 3 (LGALS3), fibronectin 1 (FN1), p-cadherin (CDH3); cell proliferation (CRIP1, IGFBP3); trans-formation (S100P, S100-A4) (Mackay et al., 2003).							
	HG1000218 N0_160000_gene_predict ion1	269	0.26		gi|6681015|r ef|NP_03178 9.1|	cysteine rich intestinal protein [Mus musculus]	
	HG1000218 N0_10000_g ene_predicti on1	196	0.36		gi|6681015|r ef|NP_03178 9.1|	cysteine rich intestinal protein [Mus musculus]	
	HG1000222 N0_1000_ge ne_ prediction1	98	0.8		gi|28603802| ref|NP_7888 40.1|	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12] [Bos taurus]	
Widely expressed.	HG1000233 N0_1000_ge	140	0.74	Nuclear; cytoplasmic.	gi|4757762|r ef|NP_00428	ring finger protein 14; androgen recep	Neurodegenerative

228

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ne_prediction1				1.1|	associated protein 54 [Homo sapiens]	hallmark is presence of ubiquitylated inclusions of insoluble protein aggregates.
	HG1000234 N0_1000_ge ne_prediction1	207	0.9		gi|4757762|ref|NP_00428 1.1|	ring finger protein 14; androgen receptor associated protein 54 [Homo sapiens]	
	HG1000234 N0_160000_gene_prediction1	207	0.9		gi|4757762|ref|NP_00428 1.1|	ring finger protein 14; androgen receptor associated protein 54 [Homo sapiens]	
	HG1000238 N0_160000_gene_prediction2	234	0.45		gi|27807167|ref|NP_7770 68.1|	anti-oxidant prote (non-selenium glutathione peroxidase, acidic calcium-independent	
	HG1000240 N0_160000_gene_prediction1	276	0.72		gi|4758756|ref|NP_00452 8.1|	nucleosome assen protein 1-like 1; HSP22-like protein interacting protein; NAP-1 related	
	HG1000245 N0_160000_gene_prediction1	793	0.28		gi|7022811|d bj|BAA9173 1.1|	unnamed protein product [Homo sapiens]	
	HG1000245 N0_5000_ge	398	0.57	FP predicted protein is	gi|7022811|d bj|BAA9173	unnamed protein product [Homo	

229

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ne_prediction1			398 aa; longer than top hit; identity 228/335 (68%) alignment; identity 83% without gaps.	1.1|	sapiens]	
	HG1000249 N0_10000_gene_predicti on1	415	0.44		gi|1449042|g b|AAB4807 0.1|	mannose-binding protein A	
	HG1000251 N0_160000_gene_predict ion1	354	0.31		gi|6005731|r ef|NP_00916 7.1|	calcium binding protein P22; SLC9A1 binding protein; calcineurin homologous protein;	
	HG1000252 N0_5000_ge ne_prediction1	139	0.8		gi|6005747|r ef|NP_00914 3.1|	ring finger protein 2 [Homo sapiens] gi|1785643|emb|CAA 71596.1| dinG [Homo sapiens]	
	HG1000254 N0_160000_gene_predict ion1	66	0.87		gi|7661790|r ef|NP_05488 6.1|	HSPC128 protein [Homo sapiens] gi|841478|gb|AAF29 092.1|AF161477_1 HSPC128 [Homo sapiens]	

230

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000262_N0_160000_gene_prediction1	478	0.92		gi|20988566|gb|AAH305 28.1|	Similar to KIAA1 protein [Homo sapiens]	
	HG1000264_N0_1000_gene_prediction1	204	0.49		gi|7661786|r ef|NP_05488 4.1|	HSPC125 protein [Homo sapiens] gi|6841472|gb|AAF29 089.1|AF161474_1 HSPC125 [Homo sapiens]	
	HG1000264_N0_1000_gene_prediction2	204	0.49		gi|7661786|r ef|NP_05488 4.1|	HSPC125 protein [Homo sapiens] gi|6841472|gb|AAF29 089.1|AF161474_1 HSPC125 [Homo sapiens]	
	HG1000270_N0_20000_gene_predicti on1	349	0.17		gi|7706353|r ef|NP_05716 3.1|	CGI-149 protein [Homo sapiens] gi|4929767|gb|AAD3 4144.1|AF151907_1 CGI-149 protein [Homo sapiens]	
	HG1000270_N0_1000_gene_prediction1	128	0.48		gi|7706353|r ef|NP_05716 3.1|	CGI-149 protein [Homo sapiens] gi|4929767|gb|AAD3 4144.1|AF151907_1 CGI-149 protein [Homo sapiens]	
	HG1000274	350	0.25		gi|28193172|	unnamed protein	

231

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	NO_160000_gene_prediction1				emb|CAD62328.1|	product [Homo sapiens]	
	HG1000276NO_160000_gene_prediction1	232	0.3		gi|8923930|ref|NP_060934.1|	uncharacterized hematopoietic stem/progenitor cells protein MDS029 [Homo sapiens]	
	HG1000276NO_5000_gene_prediction1	213	0.33		gi|8923930|ref|NP_060934.1|	uncharacterized hematopoietic stem/progenitor cells protein MDS029 [Homo sapiens]	
	HG1000278NO_5000_gene_prediction1	127	0.25		gi|8924090|ref|NP_060979.1|	hypothetical protein PRO1855 [Homo sapiens]	
	HG1000280NO_160000_gene_prediction1	494	0.63		gi|28514322|ref|XP_109734.2|	RIKEN cDNA 4732407F15 gene [Mus musculus]	
	HG1000280NO_1000_gene_prediction1	423	0.73		gi|28514322|ref|XP_109734.2|	RIKEN cDNA 4732407F15 gene [Mus musculus]	
	HG1000280NO_160000_gene_predict	494	0.63		gi|28514322|ref|XP_109734.2|	RIKEN cDNA 4732407F15 gene [Mus musculus]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ion2						
	HG1000280N0_1000_gene_prediction2	423	0.73		gi|28514322|ref|XP_1097 34.2|	RIKEN cDNA 4732407F15 gene [Mus musculus]	
	HG1000305N0_5000_gene_prediction1	232	0.53		gi|9967133|dbj|BAB1226 8.1|	hypothetical protein [Macaca fascicularis]	
	HG1000305N0_5000_gene_prediction2	232	0.53		gi|9967133|dbj|BAB1226 8.1|	hypothetical protein [Macaca fascicularis]	
	HG1000307N0_160000_gene_prediction1	173	0.75		gi|20455192|sp|Q9UKK9|NUD5_HU MAN	ADP-sugar pyrophosphatase YSA1H (Nucleoside diphosphate-linked moiety X motif 5)	
	HG1000334N0_160000_gene_prediction1	184	0.12		gi|16501181|emb|CAD10 078.1| SI:dZ150F1 3.2	(novel protein sim to human splicing factor 3a, subunit 1, 120kD (SF3A1))	
	HG1000335N0_160000_gene_prediction1	184	0.12		gi|16501181|emb|CAD10 078.1| SI:dZ150F1 3.2	(novel protein sim to human splicing factor 3a, subunit 1, 120kD (SF3A1))	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000337 N0_5000_ge ne_ prediction1	553	0.2		gi|27706766| ref|XP_2175 73.1|	ribosomal protein X-linked [Rattus norvegicus]	
	HG1000372 N0_160000_ gene_predict ion1	246	0.13		gi|21294536| gb|EAA0668 1.1|	agCP13136 [Anopheles gambiae str. PEST]	
	HG1000397 N0_5000_ge ne_ prediction1	139	0.12		gi|7705724|r ef|NP_05704 1.1|	CGI-29 protein [Homo sapiens] gi|4680697|gb|AAD2 7738.1|AF132963_1 CGI-29 protein [Homo sapiens]	
	HG1000414 N0_160000_ gene_predict ion1	487	0.49		gi|2072977|g b|AAC5127 9.1|	putative p150 [Homo sapiens]	
	HG1000439 N0_160000_ gene_predict ion1	110	0.89	Intracellular.	gi|3914909|s p|P79103|RS 4_BOVIN	40S ribosomal protein S4 gi|432939|dbj|BAA2 1078.1| ribosomal protein S4 [Bos taurus]	Turner syndrome.
	HG1000449 N0_20000_g ene_predicti on1	143	0.84		gi|27501128| ref|XP_2104 55.1|	similar to CG7874-PA [Drosophila melanogaster] [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000461 N0_160000_ gene_predict ion1	372	0.05		gi|21732487| emb|CAD38 600.1|	hypothetical protein [Homo sapiens]	
	HG1000476 N0_160000_ gene_predict ion1	103	0.17		gi|8100498|g b|AAF72329 .1|AF151371 _1 gi|8100500|gb|AAF72 330.1|AF151372_1	Ca2+-ATPase [Toxoplasma gondii] Ca2+-ATPase [Toxoplasma gondii]	
	HG1000530 N0_160000_ gene_predict ion1	366	0.08		gi|21297352| gb|EAA0949 7.1|	ebiP3838 [Anopheles gambiae str. PEST]	
	HG1000556 N0_160000_ gene_predict ion2	319	0.9		gi|106322|pi r||B34087	hypothetical protein (L1H 3' region) - human	
	HG1000584 N0_160000_ gene_predict ion1	151	0.13		gi|25051624| ref|XP_1964 90.1|	hypothetical protein XP_196490 [Mus musculus]	
	HG1000587 N0_160000_ gene_predict ion1	137	0.1		gi|18027742| gb|AAL5583 2.1|AF31832 5_1	unknown [Homo sapiens]	
	HG1000594 N0_160000	77				no_blastp_hit	

235

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	gene_prediction1						
	HG1000594 N0_160000_ gene_predict ion2	77				no_blastp_hit	
	HG1000620 N0_160000_ gene_predict ion1	310	0.19		gi|17455045| ref|XP_0630 08.1|	similar to KIAA1074 protein [Homo sapiens]	
	HG1000631 N0_40000_g ene_predicti on1	804	0.68		gi|11494381| gb|AAG357 90.1|AF2887 38_5	truncated epidermal growth factor recej [Homo sapiens]	
	HG1000686 N0_160000_ gene_predict ion1	501	0.37		gi|5070621|g b|AAD3921 4.1|AF14885 6_1	unknown [Homo sapiens]	
	HG1000712 N0_160000_ gene_predict ion1	180	0.09		gi|7959817|g b|AAF71079 .1|AF116721 _59	PRO1412 [Homo sapiens]	
	HG1000727 N0_160000_ gene_predict ion1	204	0.14		gi|26335645| dbj|BAC315 23.1|	unnamed protein product [Mus musculus]	
	HG1000743 N0 160000	310	1		gi|4505425|r ef|NP_00250	neuro-oncological ventral antigen 1,	

236

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	gene_predict ion2				6.1|	isoform 1; Neurooncological ventral antigen 1;	
	HG1000767 N0_5000_ge ne_ prediction1	553	0.2		gi|27706766| ref|XP_2175 73.1|	ribosomal protein X-linked [Rattus norvegicus]	
	HG1000822 N0_160000_ gene_predict ion1	92	0.72		gi|20883730| ref|XP_1233 11.1|	similar to histone deacetylase 1 [Mus musculus]	
	HG1000829 N0_160000_ gene_predict ion1	902	0.71		gi|6539606|g b|AAF15947 .1|	metastasis suppres protein [Homo sapiens]	
Testis.	HG1000860 N0_160000_ gene_predict ion1	610	0.83	Cell surface.	gi|7512247|p ir||I65253	disintegrin-like testicular metalloproteinase (3.4.24.-) IVb - crab-eating	Testicular cancer.
	HG1000898 N0_10000_g ene_predicti on1	236	0.72		gi|27462211| gb|AAO153 82.1|AF3343 86_1	multiple hat domains [Homo sapiens]	
	HG1000898 N0_160000_ gene_predict ion1	509	0.33		gi|27462211| gb|AAO153 82.1|AF3343 86_1	multiple hat domains [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000898N0_20000_gene_prediction1	377	0.45		gi|27462211|gb|AAO15382.1|AF334386_1	multiple hat domains [Homo sapiens]	
	HG1000902N0_160000_gene_prediction1	176	0.89		gi|14517632|dbj|BAB61032.1|	acute morphine dependence related protein 2 [Homo sapiens]	
	HG1000906N0_20000_gene_prediction1	887	0.23		gi|26346114|dbj|BAC36708.1|	unnamed protein product [Mus musculus]	
	HG1000906N0_160000_gene_predictionl	1047	0.2		gi|4505843|ref|NP_003619.1|	plakophilin 4 [Homo sapiens] gi|20139104|sp|Q99569|PKP4_HUMAN Plakophilin 4 (p0071)	
	HG1000921N0_5000_gene_prediction1	347	0.79		gi|400631|sp|P31662|NTT4_RAT	Orphan sodium- and chloride-dependent neurotransmitter transporter NTT4	
	HG1000938N0_10000_gene_predictionl	456	0.53		gi|26339054|dbj|BAC33198.1|	unnamed protein product [Mus musculus]	
	HG1000952N0_160000_gene_predict	329	0.49		gi|106322|pir||B34087	hypothetical protein (L1H 3' region) - human	

238

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ion1						
	HG1000961 N0_160000_ gene_predict ion1	257	0.13		gi|27478554| ref|XP_2098 15.1|	similar to Dumpy : shorter than wild-type protein 19 [Caenorhabditis elegans] [Homo	Cancer; cell growth; cell differentiation; cell activation; cell motility; invasion; metastasis; cell adhesion; wound healing; tissue repair; receptor tyrosine kinase; cytokine receptor.
	HG1000961 N0_160000_ gene_predict ion2	178	0.43		gi|22209001| ref|NP_6637 48.1|	tigger transposable element derived 1; jerky (mouse) homolog-like [Homo sapiens]	
	HG1001000 N0_160000_ gene_predict ion2	1798	0.3		gi|27731817| ref|XP_2187 20.1|	similar to rjs [Mus musculus] [Rattus norvegicus]	
	HG1001003 N0_160000_ gene_predict ion1	188	0.37		gi|2829147|g b|AAC0049 6.1|	lymphocyte-specific protein 1 [Homo sapiens]	
	HG1001007 N0_160000_ gene_predict ion1	459	0.5	Extra-cellular; cell surface.	gi|26334641| dbj|BAC310 21.1|	unnamed protein product [Mus musculus]	Natural product synthesis.
Skin.	HG1001009	311	0.7		gi|12698117|	hypothetical prote	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_0_gene_ prediction1				dbj|BAB218 85.1|	[Macaca fascicularis]	Defects in drd5 cause blepharospasm, a primary focal dystonia affecting the orbicularis oculi muscles; symptoms include eye irritation and frequent blinking, progressing to involuntary spasms of eyelid closure; severe cases can lead to functional blindness.
Neurons in brain limbic regions; 10-fold higher affinity for dopamine than the D1 subtype; related pseudogenes reside on chromosomes 1 & 2.	HG1001014 N0_160000_ gene_predict ion2	487	0.74	Integral membrane protein.	gi|4503391|r ef|NP_00078 9.1|	dopamine receptor D5; dopamine receptor D1B; D1beta dopamine receptor [Homo sapiens]	
	HG1001017 N0_40000_g ene_predicti on1	127	0.47		gi|25021133| ref|XP_2077 37.1|	hypothetical protein XP_207737 [Mus musculus]	
	HG1001017 N0_20000_g ene_predicti on1	121	0.49		gi|25021133| ref|XP_2077 37.1|	hypothetical protein XP_207737 [Mus musculus]	
	HG1001144 N0_160000_ gene_predict ion1	311	0.96		gi|5070621|g b|AAD3921 4.1|AF14885 6 1	unknown [Homo sapiens]	
	HG1001172 N0_160000_ gene_predict ion2	408	0.45		gi|26340706| dbj|BAC340 15.1|	unnamed protein product [Mus musculus]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1001253 N0_160000_ gene_predict ion1	195	0.85		gi|27485468| ref|XP_1706 82.2|	similar to transient receptor potential cation channel, subfamily C, member 2;	
	HG1001253 N0_160000_ gene_predict ion2	195	0.85		gi|27485468| ref|XP_1706 82.2|	similar to transient receptor potential cation channel, subfamily C, member 2;	
	HG1001267 N0_160000_ gene_predict ion1	234	0.87		gi|106322|pi r|B34087	hypothetical protein (L1H 3' region) - human	
	HG1001289 N0_160000_ gene_predict ion1	267	0.49		gi|5453998|r ef|NP_00638 2.1|	importin 7; RAN-binding protein 7 [Homo sapiens]	
	HG1001343 N0_10000_g ene_predicti on1	399	0.42		gi|4758924|r ef|NP_00456 1.1|	phosphoinositide-3-kinase, class 2, gamma polypeptide; phosphatidylinositol 3-kinase C2	
	HG1001343 N0_160000_ gene_predict ion1	454	0.37		gi|4758924|r ef|NP_00456 1.1|	phosphoinositide-3-kinase, class 2, gamma polypeptide; phosphatidylinositol 3-kinase C2	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1001390N0_160000_gene_prediction1	92	0.3		gi|914961|dbj|BAA08646.1|	Ash-s [Rattus norvegicus]	
	HG1001508N0_160000_gene_prediction2	195	0.17		gi|18557931|ref|XP_087525.1|	similar to PRO1546 [Homo sapiens]	
	HG1000084N0_160000_gene_prediction1	294	0.09		gi|25019766|ref|XP_207706.1|	hypothetical protein XP_207706 [Mus musculus]	
	HG1000084N0_160000_gene_prediction2	253	0.1		gi|25019766|ref|XP_207706.1|	hypothetical protein XP_207706 [Mus musculus]	
	HG1000209N0_160000_gene_prediction1	817	0.29		gi|11065728|emb|CAC14427.1|	dJ493F7.2 (PTD013 similar to CGI-24 protein) [Homo sapiens]	
	HG1000005N0_160000_gene_prediction1	1148	0.72		gi|5052951|gb|AAD38785.1|AF149422 2	unknown [Homo sapiens]	
High expression in lung adeno-carcinoma.	HG1000014N0_160000_gene_prediction1	270	0.92	STM.	gi|4502281|ref|NP_001670.1|	ATPase, Na+/K+ transporting, beta 3 polypeptide [Homo sapiens]	Cancer target.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000015N0_160000_gene_prediction1	1940	0.21		gi|15620899|dbj|BAB678 13.1|	KIAA1920 protein [Homo sapiens]	
	HG1000015N0_20000_gene_prediction1	980	0.43		gi|15620899|dbj|BAB678 13.1|	KIAA1920 protein [Homo sapiens]	
	HG1000015N0_5000_gene_prediction1	944	0.44		gi|15620899|dbj|BAB678 13.1|	KIAA1920 protein [Homo sapiens]	
	HG1000015N0_160000_gene_prediction2	2990	0.18		gi|20467423|ref|NP_6205 70.1|	chondroitin sulfate proteoglycan 4 [Mus musculus]	
	HG1000020N0_160000_gene_prediction1	276	0.4		gi|7263959|e mb|CAB816 42.1| bA395L14.5	(novel phosphoglucomutase like protein) [Homo sapiens]	
	HG1000020N0_5000_gene_prediction2	214	0.22		gi|7263959|e mb|CAB816 42.1| bA395L14.5	(novel phosphoglucomutase like protein) [Homo sapiens]	
	HG1000024N0_10000_gene_prediction1	295	0.31		gi|12853786|dbj|BAB298 48.1|	unnamed protein product [Mus musculus]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000026 N0_160000_ gene_predict ion1	379	0.53		gi|12248755| dbj|BAB202 65.1|	mono ATP-binding cassette protein [Homo sapiens]	
Expressed in most tissue; highest expression in brain.	HG1000030 N0_160000_ gene_predict ion1	204	0.5	Cytoplasmic	gi|16306496| ref|NP_3874 48.1|	F-box and WD-40 domain protein 1B, isoform B; F-box protein Fbw1b; beta-transducin	Neuronal related disease.
	HG1000039 N0_160000_ gene_predict ion1	105	0.14		gi|27682955| ref|XP_2440 91.1|	hypothetical protein XP_244091 [Rattus norvegicus]	
	HG1000041 N0_5000_ge ne_ prediction1	450	0.52		gi|27659604| ref|XP_2265 63.1|	similar to tripartite motif protein 9, isoform 1; homolog of rat RING finger Spring [Homo	
	HG1000043 N0_160000_ gene_predict ion1	477	0.75		gi|4503391|r ef|NP_00078 9.1|	dopamine receptor D5; dopamine receptor D1B; D1beta dopamine receptor [Homo sapiens]	
	HG1000043 N0_5000_ge ne_ prediction1	477	0.75		gi|4503391|r ef|NP_00078 9.1|	dopamine receptor D5; dopamine receptor D1B; D1beta dopamine receptor [Homo sapiens]	

244

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000044 N0_20000_gene_predicti on1	252	0.64		gi|6678990|r ef|NP_03268 7.1|	myosin Vb [Mus musculus] gi|18316|sp|P21271| MY5B_MOUSE Myosin Vb (Myosin 5B)	
	HG1000052 N0_160000_ gene_predict ion2	254	0.2		gi|7459582|p ir||S72481	probable transposase - human transposon MER37	
	HG1000052 N0_10000_g ene_predicti on1	206	0.25		gi|7459582|p ir||S72481	probable transposase - human transposon MER37	
	HG1000052 N0_20000_g ene_predicti on1	201	0.26		gi|7459582|p ir||S72481	probable transposase - human transposon MER37	
	HG1000058 N0_10000_g ene_predicti on1	390	0.38		gi|225047|pr f||1207289A	reverse transcriptase related protein	
	HG1000061 N0_5000_ge ne_ prediction1	128	0.35		gi|20910732| ref|XP_1374 48.1|	similar to Actin-like protein 2 (Actin-related protein 2) [Mus musculus]	
High expression in liver, kidney,	HG1000065 N0_5000_ge	178	0.55	Intracellular (cytoplasm).	gi|13378141| ref|NP_0547	DKFZP586A0522 protein [Homo	Cancer, metabolic disease.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
pancreas colon-ic adenocar-cinoma, cancers of pancreas, kidney, liver, bladder.	ne_prediction1				52.1|	sapiens] gi|10435697|dbj|BAB14643.1| unnamed protein product [Homo sapiens]	
	HG1000065N0_10000_gene_prediction1	178	0.55		gi|13378141|ref|NP_0547 52.1|	DKFZP586A0522 protein [Homo sapiens] gi|10435697|dbj|BAB14643.1| unnamed protein product [Homo sapiens]	
	HG1000065N0_160000_gene_prediction1	197	0.49		gi|13378141|ref|NP_0547 52.1|	DKFZP586A0522 protein [Homo sapiens] gi|10435697|dbj|BAB14643.1| unnamed protein product [Homo sapiens]	
	HG1000068N0_160000_gene_prediction1	77				no_blastp_hit	
High in kidney; expressed in all tissues examined.	HG1000070N0_0_gene_prediction1	143	0.73	STM.	gi|18203885|gb|AAH217 00.1|AAH21700	Unknown (protein for IMAGE:4052080) [Homo sapiens]	Cancer.
Overexpressed	HG1000073	268	0.3	Cytoplasmic	gi|12803571|	serine/threonine	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
in Burkitt's lymphoma; ubiquitously expressed at low levels.	N0_20000_gene_predicti on1				gb|AAH026 18.1|AAH02 618	kinase 16 [Homo sapiens]	cancers; may activate MAPK.
	HG1000075 N0_160000_gene_predict ion1	608	0.49		gi|2072972|g b|AAC5127 6.1|	putative p150 [Homo sapiens]	
High in retinal fovea, brain, heart, placenta, lung, liver, skeletal muscle, kidney.	HG1000076 N0_160000_gene_predict ion1	72	0.91	Phosphoryla se kinase, delta.	gi|115526|sp| P21251|CAL M_STIJA	Calmodulin gi|71669|pir||MCSFC U calmodulin - sea cucumber (Stichopus	
	HG1000081 N0_160000_gene_predict ion1	361	0.44		gi|20177936| sp|Q9GKX8| HS9B_HOR SE	Heat shock protein HSP 90-beta (HSP 84)	
	HG1000106 N0_160000_gene_predict ion1	282	0.17		gi|26348781| dbj|BAC380 30.1|	unnamed protein product [Mus musculus]	
	HG1000107 N0_160000_gene_predict ion1	662	0.83		gi|6708502|g b|AAD0945 4.2|	superfast myosin heavy chain [Felis catus]	
	HG1000109	98	0.91		gi|4826948|r	protein kinase, X-	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_0_gene_prediction1				ef|NP_00503 5.1|	linked [Homo sapiens] gi|1709648|sp|P51817| PKX1_HUMAN Protein kinase PKX1	
	HG1000112 N0_160000_gene_predict ion1	554	0.36		gi|106322|pi r||B34087	hypothetical protein (L1H 3' region) - human	
	HG1000116 N0_160000_gene_predict ion1	422	0.49		gi|72222|pir|| HHHU84	heat shock protein 90-beta [validated] - human gi|306891|gb|AAA36 025.1| 90kDa heat shock protein	
	HG1000126 N0_160000_gene_predict ion1	187	0.35		gi|10092689| ref|NP_0651 99.1|	hypothetical protein dJ1220B.2 [Homo sapiens]	
	HG1000130 N0_160000_gene_predict ion1	173	0.66		gi|11321591| ref|NP_0021 20.1|	high-mobility group box 2; high-mobility group (nonhistone chromosomal) prot 2	
Predominantly expressed in muscle, brain (hypothalamus).	HG1000132 N0_160000_gene_predict ion1	625	0.41	Golgi.	gi|26347765| dbj|BAC375 31.1|	unnamed protein product [Mus musculus]	RNA binding protein with RRM motif; spinocerebellar ataxia type 2 (SCA2) caused by expansion of a

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000133 N0_160000_ gene_predict ion1	121	0.82		gi|12963687| ref|NP_0759 53.1|	RIKEN cDNA 1110003H09 [Mus musculus] gi|27706342|ref|XP_2 31052.1| similar to RIKEN cDNA 1110003H09 [Mus	repeat in ataxin-2, the SCA2 gene product; expression of mutated ataxin-2 results in loss of colocalization with golgi markers (Huynh et al., 2003); ataxin-2 binding protein contributes to biological activity.
	HG1000134 N0_20000_g ene_predicti on1	371	0.37		gi|27479840| ref|XP_2082 77.1|	similar to heterogeneous nuc ribonucleoprotein C (C1/C2) [Homo sapiens]	
	HG1000134 N0_20000_g ene_predicti on2	371	0.37		gi|27479840| ref|XP_2082 77.1|	similar to heterogeneous nuc ribonucleoprotein C (C1/C2) [Homo sapiens]	
	HG1000142 N0_160000_ gene_predict	183	0.64		gi|18314408| gb|AAH219 83.1|AAH21	nucleophosmin (nucleolar phosphoprotein B2	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ion1				983	numatrin) [Homo sapiens]	
	HG1000144N0_20000_gene_prediction1	157	0.37		gi|1362931|pir||S55915	ribosomal protein 1 - human gi|550019|gb|AAA85657.1| ribosomal protein L28	
	HG1000145N0_160000_gene_prediction1	127	0.77		gi|13904866|ref|NP_00982.2|	ribosomal protein L28; 60S ribosomal protein L28 [Homo sapiens]	
	HG1000146N0_160000_gene_prediction1	177	0.81		gi|13904870|ref|NP_001000.2|	ribosomal protein 40S ribosomal protein S5 [Homo sapiens]	
	HG1000150N0_10000_gene_prediction1	129	0.91		gi|11037798|ref|NP_067621.1|	dynactin 5; dynactin 4; p25 dynactin subunit [Mus musculus]	
	HG1000152N0_160000_gene_prediction1	291	0.92		gi|27481416|ref|XP_087026.2|	similar to GIG18 [Mus musculus] [Homo sapiens]	
	HG1000161N0_160000_gene_prediction1	174	0.6		gi|24308386|ref|NP_443169.1|	similar to hypothetical protein FLJ10883 [Homo sapiens]	
High	HG1000163	201	0.71	Nuclear.		ribosomal protein	

250

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
abundance.	N0_160000_gene_predict ion1				gi|15431295| ref|NP_1502 54.1|	L13; 60S ribosomal protein L13; breast basic conserved protein 1 [Homo	
	HG1000164 N0_5000_ge ne_ prediction1	296	0.48		gi|7022560|d bj|BAA9164 4.1|	unnamed protein product [Homo sapiens]	
Ubiquitous; highest levels in brain, neurons, testis, leukemic lymphocytes, more highly expressed in patient cells from acute leukemia of different sub-types than in normal peri-pheral blood lymphocytes, nonleukemic proliferating lymphoid cells, bone marrow cells, or cells from patients	HG1000165 N0_1000_ge ne_ prediction1	45	0.75	Cytoplasmic	gi|27688981| ref|XP_2208 74.1|	similar to phosphoprotein [Rattus norvegicus]	Inhibiting phosphorylation may block proliferation of leukemic leukocytes.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
with chronic lymphoid or myeloid leukemia.							
	HG1000166 N0_160000_gene_predict ion2	907	0.94		gi|14917097| sp|Q92622|Y 226_HUMA N	Hypothetical protein KIAA0226 gi|6634001|dbj|BAA1 3215.2| KIAA0226 protein [Homo sapiens]	
	HG1000167 N0_160000_gene_predict ion1	1154	0.58		gi|22538470| ref|NP_0054 33.2|	eomesodermin; t t brain, 2 [Homo sapiens]	
	HG1000171 N0_40000_g ene_predicti on1	232	0.08		gi|17448953| ref|XP_0697 92.1|	similar to LYRIC [Rattus norvegicus] [Homo sapiens]	
	HG1000171 N0_160000_gene_predict ion1	232	0.08		gi|17448953| ref|XP_0697 92.1|	similar to LYRIC [Rattus norvegicus] [Homo sapiens]	
	HG1000175 N0_160000_gene_predict ion2	151	0.17		gi|23102869| gb|ZP_0008 9366.1|	hypothetical protein [Azotobacter vinelandii]	
Overexpressed in breast cancer	HG1000176 N0_1000 ge	767	0.51	Nuclear.	gi|19718738| ref|NP 1506	KRAB zinc finger protein KR18 [Ho	

252

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
cell line MCF7.	ne_prediction1				30.1|	sapiens]	
	HG1000176 N0_160000_gene_predict ion1	871	0.52		gi|19718738| ref|NP_1506 30.1|	KRAB zinc finger protein KR18 [Homo sapiens]	
	HG1000177 N0_160000_gene_predict ion1	377	0.79		gi|4758240|r ef|NP_00422 1.1|	endothelial differentiation, sphingolipid G-protein-coupled receptor, 5; S1P receptor	
	HG1000178 N0_160000_gene_predict ion1	185	0.75	Mitochondri al.	gi|22035592| ref|NP_0577 06.2|	mitochondrial ribosomal protein l isoform a [Homo sapiens]	
	HG1000178 N0_160000_gene_predict ion2	185	0.75	Mitochondri al.	gi|22035592| ref|NP_0577 06.2|	mitochondrial ribosomal protein l isoform a [Homo sapiens]	
Low expression in spleen, liver, pancreas, testis, thymus, heart, kidney; higher levels in hepato-cellular carcin-oma and pan-creatic adeno-	HG1000180 N0_1000_ge ne_ prediction1	185	0.86	Nuclear.	gi|18202440| sp|P82979|H CC1_HUM AN	Nuclear protein Hcc-1 (HSPC316) gi|13937971|gb|AAH 07099.1|AAH07099 Similar to RIKEN cDNA 1110005A23	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
carcinoma.							
	HG1000181 N0_10000_g ene_predicti on1	936	0.8		gi|11560152| ref|NP_0713 78.1|	zinc finger protein 335; zinc-finger/leucine-zipper co-transducer NIF1;	
	HG1000181 N0_160000_ gene_predict ion1	1615	0.83		gi|11560152| ref|NP_0713 78.1|	zinc finger protein 335; zinc-finger/leucine-zipper co-transducer NIF1;	
	HG1000183 N0_160000_ gene_predict ion1	187	0.54		gi|18148873| dbj|BAB835 17.1|	hUST3 [Homo sapiens]	
	HG1000186 N0_20000_g ene_predicti on1	21				no_blastp_hit	
	HG1000186 N0_160000_ gene_predict ion2	213	1		gi|27479871| ref|XP_2083 01.1|	similar to nonhistone chromosomal protein HMG-1 - human [Homo sapiens]	
	HG1000187 N0_20000_g ene_predicti on1	21				no_blastp_hit	
	HG1000187 N0_160000_ gene_predict	593	0.52		gi|1196433|g b|AAA8803	unknown protein	

254

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ion3				8.1|		
	HG1000189 N0_1000_ge ne_ prediction1	139	0.71		gi|20879992| ref|XP_1402 10.1|	similar to BG:DS01759.1 gene product [Drosophila melanogaster] [Mus musculus]	
	HG1000189 N0_5000_ge ne_ prediction1	233	0.36		gi|20879992| ref|XP_1402 10.1|	similar to BG:DS01759.1 gene product [Drosophila melanogaster] [Mus musculus]	
	HG1000189 N0_1000_ge ne_ prediction2	139	0.71		gi|20879992| ref|XP_1402 10.1|	similar to BG:DS01759.1 gene product [Drosophila melanogaster] [Mus musculus]	
	HG1000189 N0_5000_ge ne_ prediction2	233	0.36		gi|20879992| ref|XP_1402 10.1|	similar to BG:DS01759.1 gene product [Drosophila melanogaster] [Mus musculus]	
	HG1000195 N0_10000_g ene_predicti on1	225	0.24		gi|17390530| gb|AAH182 31.1|	Unknown (protein for MGC:19236) [Mus musculus]	
	HG1000199 N0_160000_ gene predict	354	0.84	Mitochon- dria; perox- isomes.	gi|20127408| ref|NP_0001	hydroxyacyl- Coenzyme A dehydrogenase/3-	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ion1				73.2|	ketoacyl-Coenzyme A	
	HG1000201 N0_10000_gene_predicti on1	122	0.26		gi|6680728|ref|NP_03151 0.1|	ras homolog gene family, member C; aplysia ras-related homolog 9 (RhoC); ras homolog 9	
	HG1000203 N0_5000_gene_prediction1	267	0.49		gi|26325506|dbj|BAC265 07.1|	unnamed protein product [Mus musculus]	
	HG1000204 N0_10000_gene_predicti on1	578	0.5		gi|24217433|gb|AAH386 69.1|	Unknown (protein for MGC:41921) [Homo sapiens]	
	HG1000209 N0_160000_gene_predict ion2	321	0.75		gi|22060186|ref|XP_1707 22.1|	similar to KIAA0946 protein [Homo sapiens]	
	HG1000215 N0_5000_gene_prediction1	111	0.36		gi|29135337|ref|NP_8034 89.1|	suppressor of cytokine signaling 2 [Bos taurus]	
High expression in heart, placenta, lung, kidney, prostate.	HG1000215 N0_1000_gene_prediction1	100	0.41		gi|29135337|ref|NP_8034 89.1|	suppressor of cytokine signaling 2 [Bos taurus]	
	HG1000219 N0_10000_g	128	0.78		gi|27678050|ref|XP_2256	similar to histone H2A.F/Z variant,	Anticancer target.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ene_predicti on1				55.1|	isoform 1; purine-rich binding element protein B [Homo	
	HG1000221 N0_160000_ gene_predict ion1	3776	0.54		gi|15824727| gb|AAL0945 9.1|AF31769 6_1	macrophin 1 isoform 4 [Homo sapiens]	
	HG1000221 N0_20000_g ene_predicti on1	3648	0.52		gi|15824727| gb|AAL0945 9.1|AF31769 6_1	macrophin 1 isoform 4 [Homo sapiens]	
	HG1000223 N0_160000_ gene_predict ion1	132	0.67		gi|20380930| gb|AAH282 27.1|	Similar to hypothetical protein PRO2221 [Homo sapiens]	
	HG1000225 N0_160000_ gene_predict ion1	82	0.87		gi|28071124| emb|CAD61 943.1|	unnamed protein product [Homo sapiens]	
	HG1000235 N0_160000_ gene_predict ion1	40	0.75		gi|27669524| ref|XP_2203 21.1|	similar to ATPase, H+ transporting, lysosomal 13kD, V1 subunit G isoform 1; ATPase, H+	
	HG1000236 N0_160000_ gene_predict ion1	708	0.68		gi|25020161| ref|XP_2075 71.1|	similar to hypothetical protein FLJ35630 [Homo sapiens] [Mus	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
						musculus]	
	HG1000238 N0_160000_ gene_predict ion1	180	0.54		gi|27807167| ref|NP_7770 68.1|	anti-oxidant prote (non-selenium glutathione peroxidase, acidic calcium-independent	
	HG1000238 N0_5000_ge ne_ prediction1	141	0.69		gi|27807167| ref|NP_7770 68.1|	anti-oxidant prote (non-selenium glutathione peroxidase, acidic calcium-independent	
	HG1000239 N0_160000_ gene_predict ion1	264	0.69		gi|4758756|r ef|NP_00452 8.1|	nucleosome asser protein 1-like 1; HSP22-like protein interacting protein; NAP-1 related	
	HG1000241 N0_160000_ gene_predict ion1	75	0.46		gi|26337731| dbj|BAC325 51.1|	unnamed protein product [Mus musculus]	
	HG1000243 N0_160000_ gene_predict ion1	90	1		gi|5031749|r ef|NP_00550 8.1|	high-mobility group nucleosomal binding domain 2; nonhistone chromosomal protein	
	HG1000243 N0_160000_ gene_predict ion2	299	0.29		gi|18555712| ref|XP_0961 98.1|	hypothetical protein XP_096198 [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000245 N0_1000_gene_prediction1	286	0.79		gi|7022811|dbj|BAA9173 1.1|	unnamed protein product [Homo sapiens]	
	HG1000250 N0_160000_gene_predict ion1	528	0.85	STM (based on bioinformatics); mito-chondrial matrix (based on genecard); hydroxyacyl-Coenzyme A dehydrog-enase/3-ketoacyl-Coenzyme A thiolase/eno yl-Coenzyme A hydratase (trifunctiona l protein), alpha subunit.	gi|595267|gb |AAA56664. 1|	78 kDa gastrin-binding protein	
	HG1000252 N0_160000_gene_predict	70	0.92		gi|16741485| gb|AAH165 58.1|	Unknown (protein for IMAGE:3586350) [Mus musculus]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ion1						
	HG1000255 N0_10000_g ene_predicti on1	121	0.34		gi|3212110|e mb|CAA767 59.1|	prefoldin subunit 1 [Homo sapiens]	
	HG1000262 N0_160000_ gene_predict ion2	363	0.87		gi|106322|pi r||B34087	hypothetical protein (L1H 3' region) - human	
	HG1000263 N0_160000_ gene_predict ion1	134	0.49		gi|7512937|p ir||T08783	hypothetical protein DKFZp586O0120.1 - human (fragment)	
	HG1000264 N0_5000_ge ne_ prediction1	255	0.39		gi|7661786|r ef|NP_05488 4.1|	HSPC125 protein [Homo sapiens] gi|6841472|gb|AAF29 089.1|AF161474_1 HSPC125 [Homo sapiens]	
	HG1000264 N0_5000_ge ne_ prediction2	255	0.39		gi|7661786|r ef|NP_05488 4.1|	HSPC125 protein [Homo sapiens] gi|6841472|gb|AAF29 089.1|AF161474_1 HSPC125 [Homo sapiens]	
	HG1000265 N0_160000_ gene_predict	228	0.52		gi|7661886|r ef|NP_05557	DAZ associated protein 2; KIAA0058 gene product [Hon	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ion1				9.1|	sapiens]	
	HG1000266 N0_0_gene_ prediction1	143	0.93	Nuclear; LPIN3 (lipin 3) (based on genecard); 727 aa, 80,107 daltons.	gi|23821841| sp||Q9BQK8 _3	[Segment 3 of 3] Lipin 3 (Lipin 3-like)	
	HG1000266 N0_160000_ gene_predict ion1	797	0.64		gi|26340094| dbj|BAC337 10.1|	unnamed protein product [Mus musculus]	
	HG1000270 N0_160000_ gene_predict ion1	353	0.17		gi|7706353|r ef|NP_05716 3.1|	CGI-149 protein [Homo sapiens] gi|4929767|gb|AAD3 4144.1|AF151907_1 CGI-149 protein [Homo sapiens]	
	HG1000271 N0_10000_g ene_predicti on1	461	0.65		gi|27499559| ref|XP_0626 69.7|	similar to lactate dehydrogenase A - like [Homo sapiens]	
	HG1000271 N0_160000_ gene_predict ion1	615	0.49		gi|27499559| ref|XP_0626 69.7|	similar to lactate dehydrogenase A - like [Homo sapiens]	
	HG1000273 N0_160000	115	0.14		gi|22954170| gb|ZP_0000	hypothetical protein [Nitrosomonas	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	gene_prediction1				1971.1|	europaea]	
Brain	HG1000295N0_160000_gene_prediction1	229	0.69		gi|12232393|ref|NP_0735 73.1|	hypothetical protein FLJ14153 [Homo sapiens]	Small molecule target to treat depression, obesity, pain, schizophrenia.
	HG1000296N0_160000_gene_prediction1	133	0.7		gi|25054735|ref|XP_1928 39.1|	ATPas, class II, type 9B [Mus musculus]	
	HG1000299N0_160000_gene_prediction1	180	0.77		gi|4503729|r ef|NP_00200 5.1|	FK506-binding protein 4; FK506-binding protein 4 (59kD); T-cell FK506-binding protein, 59kD;	Cancer, immunosuppression.
	HG1000300N0_10000_gene_prediction1	768	0.72		gi|106322|pi r|B34087	hypothetical protein (L1H 3' region) - human	
	HG1000306N0_0_gene_prediction1	30				no_blastp_hit	
	HG1000306N0_0_gene_prediction2	30				no_blastp_hit	
	HG1000312N0_160000	351	0.27	ER; nuclear; localization	gi|4506283|r ef|NP_00345	protein tyrosine phosphatase type	PRL-1, one of a family of 3

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	gene_predict ion1			is dependent on farnesyl-ation and correllates with cell cycle.	4.1|	IVA, member 1; Protein tyrosine phosphatase IVA1 [Homo	by a c-terminal prenylation motif, is associated with cell proliferation and in regenerating liver, mit- pancreatic cancers; various cancer cell lines; antisense to PRL-1 may induce apoptosis in certain cells; prl-1 tx cells have short-ened doubling times; dom-inant negative delays pro-gression through M phase of cell cycle; involved in mitotic spindle function (Wang et al., 2002); other family members also asso-ciated with cancer; prl-2 with prostate cancer, prl-3 with colon metastasis (Wang et al., 2002; Saha et al., 2001).
	HG1000314 N0_1000_ge ne_ prediction1	85	0.91		gi|4506285|r ef|NP_00347 0.1|	protein tyrosine phosphatase type IVA, member 2, isoform 1; protein tyrosine phosphatase	
	HG1000315	149	0.87	Intracellular	gi|280853|pi	protein-tyrosine-	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_160000_gene_prediction1				r||A60345	phosphatase (EC 3.1.3.48) 11A - human	monoblastic leukemia; disorders in hematopoeisis; Noonan syndrome.
	HG1000330 N0_160000_gene_prediction2	87	0.19		gi|15809538|gb|AAK539 57.1|	MxiC [Escherichia coli] gi|15809540|gb|AAK 53958.1| MxiC [Escherichia coli]	
	HG1000330 N0_160000_gene_prediction4	133	0.17		gi|21297054|gb|EAA0919 9.1|	ebiP3674 [Anophe gambiae str. PEST]	
	HG1000337 N0_1000_gene_prediction1	171	0.89	Intracellular.	gi|27706766|ref|XP_2175 73.1|	ribosomal protein X-linked [Rattus norvegicus]	Turner syndrome.
	HG1000357 N0_20000_gene_prediction1	2992	0.99		gi|24212087|sp||Q9Y6V0 _2	[Segment 2 of 2] Piccolo protein (Aczonin)	
	HG1000358 N0_5000_gene_prediction1	407	0.88		gi|2072948|gb|AAC5126 1.1|	putative p150 [Homo sapiens]	Hemophilia.
	HG1000396 N0_160000_gene_prediction2	41				no_blastp_hit	

264

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000401 N0_10000_g ene_predicti on1	150	0.1		gi|28573982| ref|NP_7879 55.1|	CG11023-PA [Drosophila melanogaster] gi|28380291|gb|AAO 41164.1| CG11023-PA [Drosophila melanogaster]	
	HG1000414 N0_160000_ gene_predict ion2	224	0.25		gi|3513512|g b|AAC3384 7.1|	nongradient byssal precursor [Mytilus edulis]	
	HG1000416 N0_160000_ gene_predict ion1	119	0.75		gi|106322|pi r|B34087	hypothetical protein (L1H 3' region) - human	
	HG1000428 N0_160000_ gene_predict ion1	236	0.55		gi|106322|pi r|B34087	hypothetical protein (L1H 3' region) - human	
	HG1000441 N0_160000_ gene_predict ion1	82	0.2		gi|23059556| ref|ZP_0008 4515.1|	hypothetical protein [Pseudomonas fluorescens PfO-1]	
	HG1000441 N0_160000_ gene_predict ion2	309	0.91		gi|106322|pi r|B34087	hypothetical protein (L1H 3' region) - human	
	HG1000446	584	0.47		gi|2072948|g	putative p150 [Ho	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	NO_160000_gene_prediction1				b|AAC5126 1.1|	sapiens]	
	HG1000446 NO_160000_gene_prediction2	361	0.59		gi|2072953|g b|AAC5126 4.1|	putative p150 [Homo sapiens]	
Expressed in many tissues, including fetal cochlea.	HG1000449 NO_160000_gene_prediction2	528	0.64	ER.	gi|13173471| ref|NP_0769 27.1|	transmembrane protease, serine 3 isoform 1; serine protease TADG12; Transmembrane	Nerve degeneration.
	HG1000451 NO_160000_gene_prediction1	181	1		gi|17434360| ref|XP_0652 19.1|	similar to RNA-binding protein LIN-28 [Homo sapiens]	
	HG1000455 NO_10000_gene_prediction1	421	0.08		gi|20824899| ref|XP_1442 55.1|	hypothetical protein XP_144255 [Mus musculus]	
	HG1000461 NO_10000_gene_prediction1	150	0.1		gi|28573982| ref|NP_7879 55.1|	CG11023-PA [Drosophila melanogaster] gi|28380291|gb|AAO 41164.1| CG11023-PA [Drosophila melanogaster]	
	HG1000476	48				no blastp hit	

266

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_1000_gene_prediction1						
	HG1000510 N0_160000_gene_predict ion1	194	0.13		gi|14249206| ref|NP_1160 44.1|	hypothetical protein MGC10997 [Homo sapiens]	Cancer; inflammation.
	HG1000513 N0_160000_gene_predict ion1	305	0.55	Cytosol.	gi|26345590| dbj|BAC364 46.1|	unnamed protein product [Mus musculus]	
	HG1000524 N0_160000_gene_predict ion1	42	0.61		gi|13959686| sp|P28476|G AR2_HUM AN	Gamma-aminobutyric-acid receptor rho-2 subı precursor (GABA(A) receptor)	
	HG1000530 N0_20000_g ene_predicti on1	132	0.5		gi|22042664| ref|XP_0674 45.5|	similar to KIAA1998 protein [Homo sapiens]	
	HG1000530 N0_160000_gene_predict ion2	95	0.24		gi|25010219| ref|NP_7346 14.1|	Unknown [Streptococcus agalactiae NEM316]	
	HG1000534 N0_160000_gene_predict ion1	150	0.6		gi|2072972|g b|AAC5127 6.1|	putative p150 [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000545 N0_160000_gene_predict ion1	346	1		gi|4885329|r ef|NP_00529 5.1|	G protein-coupled receptor 41 [Homo sapiens]	
	HG1000553 N0_160000_gene_predict ion1	256	0.44		gi|21750183| dbj|BAC037 36.1|	unnamed protein product [Homo sapiens]	
	HG1000560 N0_160000_gene_predict ion2	41				no_blastp_hit	
	HG1000566 N0_40000_g ene_predicti on1	47	0.78		gi|14249206| ref|NP_1160 44.1|	hypothetical protein MGC10997 [Homo sapiens]	
	HG1000566 N0_160000_gene_predict ion1	97	0.38		gi|14249206| ref|NP_1160 44.1|	hypothetical protein MGC10997 [Homo sapiens]	
	HG1000623 N0_160000_gene_predict ion1	83	0.22		gi|28828085| gb|AAO507 68.1|	hypothetical protein [Dictyostelium discoideum]	
	HG1000624 N0_160000_gene_predict ion1	466	0.3		gi|17485449| ref|XP_0663 97.1|	similar to ebiP4655 [Anopheles gambiae str. PEST] [Homo sapiens]	

268

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000650 N0_160000_gene_predict ion1	222	0.5		gi|22049992| ref|XP_0698 27.2|	similar to 40S RIBOSOMAL PROTEIN S17 [Homo sapiens]	
	HG1000659 N0_20000_g ene_predicti on1	402	0.17		gi|27498395| ref|XP_2106 62.1|	similar to hypothetical protein; sequence orphan, low similarity to glycoamylases and	
	HG1000661 N0_20000_g ene_predicti on1	3065	0.99		gi|24212087| spl|Q9Y6V0 _2	[Segment 2 of 2] Piccolo protein (Aczonin)	
	HG1000664 N0_160000_ gene_predict ion1	141	0.8		gi|27732755| ref|XP_2163 49.1|	similar to histone deacetylase [Mus musculus] [Rattus norvegicus]	
	HG1000690 N0_20000_g ene_predicti on1	70				no_blastp_hit	
	HG1000690 N0_20000_g ene_predicti on2	70				no_blastp_hit	
	HG1000696 N0_20000_g ene_predicti	74				no_blastp_hit	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	on1						
	HG1000696 N0_40000_g ene_predicti on1	196	0.12		gi|22086547| gb|AAM906 73.1|AF4023 16_1	surface protein SdrI [Staphylococcus saprophyticus]	
	HG1000697 N0_160000_ gene_predict ion1	226	0.14		gi|24586197| ref|NP_6102 67.1|	pawn CG11101-PA gi|10727776|gb|AAF5 9274.2| CG11101-PA [Drosophila melanogaster]	
	HG1000704 N0_160000_ gene_predict ion1	156	0.21		gi|27480400| ref|XP_2119 12.1|	hypothetical protein XP_211912 [Homo sapiens]	
	HG1000711 N0_20000_g ene_predicti on1	103	0.21		gi|16121909| ref|NP_4052 22.1|	isocitrate dehydrogenase [NADP] [Yersinia pestis]	
	HG1000740 N0_10000_g ene_predicti on1	303	0.13		gi|28921429| gb|EAA3073 5.1|	predicted protein [Neurospora crassa]	
	HG1000743 N0_160000_ gene_predict ion1	92	0.14		gi|6983861|d bj|BAA9079 6.1|	Similar to gypsy-t retrotransposon RIRE8B DNA, internal region (AB014741) [Oryza sativa	

THIS PAGE BLANK (USPTO)

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000781 N0_160000_ gene_predict ion1	347	0.57		gi|14290590| gb|AAH090 74.1|AAH09 074	Similar to CGI-62 protein [Homo sapiens]	
	HG1000781 N0_160000_ gene_predict ion2	233	0.43		gi|106322|pi r|B34087	hypothetical protein (L1H 3' region) - human	
	HG1000788 N0_1000_ge ne_ prediction1	688	0.87		gi|27662804| ref|XP_2219 54.1|	similar to slit homolog 1; Slit1 [Rattus norvegicus]	
	HG1000808 N0_160000_ gene_predict ion1	212	0.07		gi|20522131| ref|NP_6206 76.1|	movement protein [Sweet clover necr mosaic virus]	
	HG1000817 N0_160000_ gene_predict ion1	306	0.35		gi|27684119| ref|XP_2146 03.1|	similar to dnaK-type molecular chaperone hsc73 - rat [Rattus norvegicus]	
	HG1000822 N0_20000_g ene_predicti on1	175	0.35		gi|20883730| ref|XP_1233 11.1|	similar to histone deacetylase 1 [Mus musculus]	
	HG1000842 N0_160000_ gene_predict ion1	318	0.89		gi|106322|pi r|B34087	hypothetical protein (L1H 3' region) - human	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1000842N0_160000_gene_prediction2	537	0.1		gi|257757|gb|AAB23712.1|	reverse transcriptase homolog {HOMORT2, retrotransposon} [human, Peptide Transposon Partial, 114	
	HG1000870N0_160000_gene_prediction1	160	0.81		gi|27369942|ref|NP_7662 46.1|	RIKEN cDNA C330017E21; hypothetical protein 9530051F04 [Mus musculus]	
	HG1000870N0_160000_gene_prediction2	160	0.81		gi|27369942|ref|NP_7662 46.1|	RIKEN cDNA C330017E21; hypothetical protein 9530051F04 [Mus musculus]	
High expression in testis, medulla, hippocampus.	HG1000878N0_20000_gene_prediction1	249	0.3	MTM.	gi|27686461|ref|XP_2374 37.1|	similar to testis-specific transporter TST1 [Rattus norvegicus]	Cancer, neuronal disorders.
	HG1000878N0_20000_gene_prediction2	249	0.3		gi|27686461|ref|XP_2374 37.1|	similar to testis-specific transporter TST1 [Rattus norvegicus]	
High expression in frontal lobe posterior rhom-bomeres,	HG1000906N0_5000_gene_prediction1	221	0.45		gi|4505843|ref|NP_00361 9.1|	plakophilin 4 [Homo sapiens] gi|20139104|sp|Q9956 9|PKP4_HUMAN	Cancer; markers for predicting clinical outcomes; Alzheimer's

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
anterior rhombomeres, endo-metrioid ovarian metastasis, adult brain.						Plakophilin 4 (p0071)	tumors (Papagerakis et al., 2003).
	HG1000906 N0_160000_ gene_predict ion2	1066	0.62		gi|5052951|g b|AAD3878 5.1|AF14942 2_2	unknown [Homo sapiens]	
	HG1000910 N0_160000_ gene_predict ion1	214	0.55		gi|24431977| ref|NP_0605 23.2|	hypothetical protein FLJ10307 [Homo sapiens]	Transcriptional activation; cancer; modulation of cell growth, differentiation, activation.
	HG1000948 N0_160000_ gene_predict ion1	344	0.85		gi|106322|pi r||B34087	hypothetical protein (L1H 3' region) - human	Cancer; modulation of cell growth, activation, differentiation; biotechnology; gene cloning; directed evolution; genetic engineering; generation of recombinant DNA; generation of transgenic, knock-out, chimeric organisms.
	HG1000955 N0_160000_ gene_predict ion1	205	0.24		gi|12849880| dbj|BAB285 17.1|	unnamed protein product [Mus musculus]	
	HG1000959 N0_160000	525	0.24		gi|27713370| ref|XP_2323	similar to RIKEN cDNA 1110014F1	

273.

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	gene_predict ion1				71.1|	[Mus musculus] [Rattus norvegicus]	Extracellular domain, including variable part thereof, as antibody target; palmitoylation of intracellular domain; inhibitors useful for lipid modification; neutralizing antibodies useful for modulting virus entry.
	HG1000959 N0_5000_ge ne_ prediction1	406	0.31		gi|27713370| ref|XP_2323 71.1|	similar to RIKEN cDNA 1110014F12 [Mus musculus] [Rattus norvegicus]	
	HG1000990 N0_5000_ge ne_ prediction1	118	0.3		gi|10946762| ref|NP_0673 82.1|	triggering receptor expressed on myeloid cells 3; triggering receptor expressed on	
	HG1000994 N0_10000_g ene_predicti on1	70	0.67		gi|21361800| ref|NP_4780 67.2|	chromosome 21 o reading frame 63 [Homo sapiens]	Down's syndrome; Alzheimer's disease; Usher Syndrome type 1E; Amyotrophic Lateral Sclerosis; cancer; acute lymphoblastic leukemia; multiple myeloma.
	HG1000994 N0_160000_ gene_predict ion2	206	0.42		gi|27498059| ref|XP_0681 66.3|	similar to olfactory receptor MOR145-2 [Mus musculus] [Homo sapiens]	
	HG1000994 N0_10000_g ene_predicti	70	0.67		gi|21361800| ref|NP_4780 67.2|	chromosome 21 o reading frame 63 [Homo sapiens]	Down's syndrome; Alzheimer's disease; Usher

274

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	on2						Amyotrophic Lateral Sclerosis; cancer; acute lymphoblastic leukemia; multiple myeloma.
Brain (limbic regions, striatum, nucleus accumbens, olfactory tubercle, frontal cortex, cortical areas (layers II, IV, and VI), dentate gyrus, hippocampus); tumors; neurons; may complex with GABA-gated channels.	HG1001001N0_160000_gene_prediction1	477	0.75	Cell surface.	gi|4503391|ref|NP_00078 9.1|	dopamine receptor D5; dopamine receptor D1B; D1beta dopamine receptor [Homo sapiens]	Neuronal disease; blepharospasm; focal dystonia; focal torsion dystonia; primary cervical dystonia; dopamine beta hydroxylase deficiency; aromatic l-amino acid hydroxylase deficiency; monoamineoxidase deficiency; dopamine-related disorders; Parkinson's disease; depression; post-traumatic stress disorder; psycho-motor diseases; schizophrenia; cancer; angiogenesis/anti-angiogenesis.
Neurons, tumors.	HG1001001N0_0_gene_prediction1	405	0.82	Cell surface.	gi|27498059|ref|XP_0681 66.3|	similar to olfactory receptor MOR145-2 [Mus musculus] [Homo sapiens]	Biosensor, chemosensor; neuronal disorder; baroreflex failure; dopamine beta hydroxylase deficiency; tetrahydro-biopterin deficiency;

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1001002 N0_160000_gene_predict ion1	805	0.17		gi|27699752| ref|XP_2241 00.1|	similar to T-cell receptor alpha-chain variable region [Rattus norvegicus]	hydroxylase deficiency; Menkes disease; mono-amineoxidase deficiency; dopamine-Parkinson's disease; depression; post-traumatic stress disorder; emphysema; bronchitis; allergic conditions; cardiac stimulation; ocular disease; glaucoma; retinitis pigmentosa; congenital night blindness; neurodegenerative disease; cancer; pheochromocytoma; neuroblastoma; chemodectina; familial paraganglioma syndrome.
Neurons, tumors.	HG1001003 N0_0_gene_ prediction1	405	0.82	Cell surface.	gi|27498059| ref|XP_0681 66.3|	similar to olfactory receptor MOR145-2 [Mus musculus] [Homo sapiens]	Biosensor; chemosensor; neuronal disorder; baroreflex failure; dopamine beta hydroxylase deficiency; tetrahydro-biopterin deficiency;

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
							hydroxylase deficiency; Menkes disease; mono-amineoxidase deficiency; dopamine-Parkinson's disease; depression; post-traumatic stress disorder; emphysema; bronchitis; allergic conditions; cardiac stimulation; ocular disease; glaucoma; retinitis pigmentosa; congenital night blindness; neurodegenerative disease; cancer; pheochromocytoma; neuroblastoma; chemodectina; familial paraganglioma syndrome.
	HG1001007N0_160000_gene_prediction2	156	0.32		gi|27695217|gb|AAH439 49.1|	Similar to hypothetical protein FLJ12666 [Xenopus laevis]	
	HG1001011N0_160000_gene_prediction1	275	0.43		gi|22122413|ref|NP_6660 85.1|	hypothetical protein MGC31450 [Mus musculus]	
	HG1001011N0_160000_gene_predict	283	0.78		gi|25047957|ref|XP_1305 82.2|	similar to hypothetical protein MGC14161 [Hom	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	ion2					sapiens] [Mus musculus]	
Neurons in brain limbic regions; 10-fold higher affinity for dopamine than the D1 subtype; related pseudogenes reside on chro-mosomes 1 & 2.	HG1001014 N0_160000_ gene_predict ion1	477	0.75	Integral membrane protein.	gi|4503391|r ef|NP_00078 9.1|	dopamine receptor D5; dopamine receptor D1B; D1beta dopamine receptor [Homo sapiens]	Defects in drd5 cause blepharospasm, a primary focal dystonia affecting the orbicularis oculi muscles; symptoms include eye irritation and frequent blinking, progressing to involuntary spasms of eyelid closure; severe cases can lead to functional blindness.
Neurons in brain limbic regions; 10-fold higher affinity for dopamine than the D1 subtype; related pseudogenes reside on chro-mosomes 1 & 2.	HG1001014 N0_5000_ge ne_ prediction1	477	0.75	integral membrane protein	gi|4503391|r ef|NP_00078 9.1|	dopamine receptor D5; dopamine receptor D1B; D1beta dopamine receptor [Homo sapiens]	Defects in drd5 cause blepharospasm, a primary focal dystonia affecting the orbicularis oculi muscles; symptoms include eye irritation and frequent blinking, progressing to involuntary spasms of eyelid closure; severe cases can lead to functional blindness.
	HG1001017 N0_160000_ gene_predict ion1	301	0.19		gi|25021133| ref|XP_2077 37.1|	hypothetical protein XP_207737 [Mus musculus]	
	HG1001020	75	0.16		gi|22536912|	riboflavin	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	NO_160000_gene_predict ion1				ref|NP_6877 63.1|	biosynthesis protein RibA [Streptococcus agalactiae 2603V/R]	Cancer, neuronal diseases, inflmmatory diseases, immune modulation.
Human brain.	HG1001024 NO_160000_gene_predict ion1	899	0.85	TM.	gi|28522815| ref|XP_1498 40.2|	expressed sequence AW455481 [Mus musculus]	
Brain.	HG1001024 NO_160000_gene_predict ion2	899	0.85		gi|28522815| ref|XP_1498 40.2|	expressed sequence AW455481 [Mus musculus]	
	HG1001031 NO_160000_gene_predict ion1	217	0.37		gi|25071690| ref|XP_1935 91.1|	RIKEN cDNA 1110012D08 [Mus musculus] gi|22902311|gb|AAH 37624.1| Unknown (protein for MGC:47459) [Mus	
Bone marrow, brain, heart, skeletal muscle, liver, kidney, lung.	HG1001035 NO_5000_ge ne_ prediction1	79	0.53		gi|21755758| dbj|BAC047 53.1|	unnamed protein product [Homo sapiens]	
	HG1001043 NO_160000_gene_predict ion1	1136	0.24		gi|5453724|r ef|NP_00614 3.1|	lymphoid-restricted membrane protein [Homo sapiens]	
	HG1001046	90	0.24		gi|12044868|	ATP-dependent R.	

279

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	NO_5000_gene_prediction1				ref|NP_0726 78.1|	helicase, putative [Mycoplasma genitalium]	
	HG1001046 NO_160000_gene_prediction1	226	0.29		gi|27478106| ref|XP_2118 57.1|	hypothetical protein XP_211857 [Homo sapiens]	
	HG1001047 NO_1000_gene_prediction1	129	0.41	Nuclear.	gi|27667816| ref|XP_2214 77.1|	similar to eukaryotic translation initiation factor 4A1; initiation factor eIF-4A long	
	HG1001048 NO_160000_gene_prediction1	374	0.86		gi|24980856| gb|AAH398 68.1|	Unknown (protein for MGC:49034) [Homo sapiens]	
Bone marrow, lung.	HG1001048 NO_160000_gene_prediction2	92	0.79	Nuclear; importin is a cytoplasmic protein that can bind to nuclear pore complexes.	gi|938026|db j|BAA07269 .1|	Ran-binding protein 1 [Homo sapiens]	Cancer; inflammation.
	HG1001144 NO_20000_gene_predicti on1	222	0.96		gi|5070621|g b|AAD3921 4.1|AF14885 6_1	unknown [Homo sapiens]	
Ubiquitously expressed; over-	HG1001148 NO_160000	398	0.92	Type I membrane	gi|15778976| gb|AAH145	a disintegrin and metalloproteinase	Thrombosis and

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
expressed in artherosclerotic lesions; constitutively expressed in cultered endothelium, smooth muscle.	gene_prediction2			protein.	66.1|AAH14566	domain 15 (metargidin) [Homo sapiens]	00000498; mammalian epididymal apical protein 1 (EAP I) (Medline 00000515); associated with the sperm membrane and may play a role in sperm maturation; cancer, inflammation, fertility.
	HG1001172N0_160000_gene_prediction1	604	0.73		gi|27676974|ref|XP_2237 14.1|	similar to hypothetical protein MGC38936 [Mus musculus] [Rattus norvegicus]	
	HG1001172N0_20000_gene_prediction1	399	0.46		gi|26340706|dbj|BAC340 15.1|	unnamed protein product [Mus musculus]	
	HG1001194N0_160000_gene_prediction1	219	0.12		gi|22999550|gb|ZP_0004 3524.1|	hypothetical protein [Magnetococcus sp. MC-1]	
Brain, heart, kidney, lung.	HG1001223N0_160000_gene_prediction1	476	0.95		gi|5070622|gb|AAD3921 5.1|AF14885 6_2	unknown [Homo sapiens]	
	HG1001284N0_160000_gene_prediction1	685	1		gi|27485931|ref|XP_0847 36.6|	similar to CG5815-PA [Drosophila melanogaster] [Homo sapiens]	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	HG1001284 N0_160000_gene_prediction2	519	1		gi|27485931|ref|XP_084736.6|	similar to CG5815-PA [Drosophila melanogaster] [Homo sapiens]	
Ubiquitous.	HG1001292 N0_160000_gene_prediction1	151	0.92	Intracellular.	gi|4502317|ref|NP_001687.1|	ATPase, H+ transporting, lysosomal 31kD, V1 subunit E isoform 1; V-ATPase, subunit E;	
	HG1001302 N0_160000_gene_prediction1	276	0.21		gi|14149671|ref|NP_055927.1|	microtubule associated testis specific serine/threonine protein kinase; KIAA0807 protein [Homo	
	HG1001323 N0_160000_gene_prediction1	99	0.14		gi|17544358|ref|NP_502905.1|	Predicted CDS, neurotransmitter-gated ion-channel with standard ligand binding domain but	
	HG1001328 N0_5000_gene_prediction1	200	0.15		gi|7480970|pir||T36972	probable membrane associated protein - Streptomyces coelicolor (fragment)	
	HG1001328 N0_40000_gene_prediction1	666	0.09		gi|225047|prf||1207289A	reverse transcriptase related protein	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
Brain, pancreatic cancer, substantia nigra, frontal lobe, pituitary adenomas.	HG1001331 N0_0_gene_prediction1	296	0.41	MTM (type I).	gi|20381292| gb|AAH277 70.1|	stromal cell derived factor receptor 2 [Mus musculus]	Neuronal disease, inflammatory liver disorders; stromal cell-derived factor-1 from biliary epithelial cells recruits CXCR4-positive cells (Terada et al., 2003); similar to human SDR2 cDNA clone (100% ID with nucleotide sequence, 118/145 (81%) with aa sequence).
	HG1001348 N0_160000_gene_predict ion1	184	0.65		gi|20846538| ref|XP_1500 33.1|	similar to hypothetical protein XP_150033 [Mus musculus]	
	HG1001349 N0_160000_gene_predict ion1	236	0.19		gi|20831940| ref|XP_1442 79.1|	hypothetical protein XP_144279 [Mus musculus]	
	HG1001354 N0_160000_gene_predict ion1	510	1		gi|20542050| ref|XP_0333 70.6|	similar to KIAA1762 protein [Homo sapiens]	
	HG1001361 N0_160000_gene_predict ion1	137	0.78		gi|20345901| ref|XP_1098 24.1|	similar to hypothetical protein XP_109824 [Mus musculus]	
	HG1001376 N0_160000	455	0.99		gi|7512821|p ir||T00347	hypothetical protein DKFZp566G1246.	

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public Protein	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	gene_predict ion1					version I - human (fragment)	
	HG1001376 N0_5000_ge ne_ prediction1	224	0.99		gi|7512821|p ir||T00347	hypothetical protein DKFZp566G1246.1, version I - human (fragment)	
	HG1001376 N0_20000_g ene_predicti on1	625	0.99		gi|7512821|p ir||T00347	hypothetical protein DKFZp566G1246.1, version I - human (fragment)	
	HG1001376 N0_5000_ge ne_ prediction2	224	0.99		gi|7512821|p ir||T00347	hypothetical protein DKFZp566G1246.1, version I - human (fragment)	
	HG1001376 N0_5000_ge ne_ prediction3	296	0.75		gi|22052701| ref|XP_0519 56.8|	similar to hypothetical protein DKFZp566G1246.1, version I - human (fragment) [Homo no_blastp_hit	
	HG1001436 N0_5000_ge ne_ prediction1	58					
	HG1001436 N0_20000_g ene_predicti on1	170	0.24		gi|22982273| ref|ZP_0002 7554.1|	hypothetical protein [Burkholderia fungorum]	
	HG1001436	195	0.12		gi|16127778|	hypothetical prote	

284

Expression Profile	FP ID	Length, Predicted Protein	Covered by Public	Subcellular Localization	Top Hit Accession No.	Top Hit Annotation	Utility
	N0_160000_gene_prediction1				ref|NP_4223 42.1|	[Caulobacter crescentus CB15]	Endometriosis.
	HG1001484 N0_160000_gene_prediction1	371	0.47		gi|106322|pir||B34087	hypothetical protein (L1H 3' region) - human	
	HG1001500 N0_160000_gene_prediction1	203	0.86		gi|112077|pir||A36089	nucleolar phosphoprotein B2 - rat	
Normal endo-metrium, med-ulloblastoma, whole embryo.	HG1001500 N0_160000_gene_prediction2	176	0.66		gi|27484907| ref|XP_2103 58.1|	similar to p40 [Homo sapiens]	
	HG1001508 N0_160000_gene_prediction1	73	0.26		gi|23134311| ref|ZP_0011 6085.1|	hypothetical protein [Synechococcus sp. WH 8102]	
	HG1000898 N0_5000_gene_predictio n1	331			AF334386_1	multiple hat domains [Homo sapiens]	

285

Table 2. Characteristics of the Claimed Sequences, and of the HumanProtein With the Highest Degree of Similarity to Each

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG100021 4N 0_160000_ge ne_prediction 1	99	ig		1 (1-18)		0	gi|2114236|dbj|BAA19998.1|	V5-4 [Homo sapiens]
HG100032 3N 0_160000_ge ne_prediction 1	475	PLAT lipase		1 (1-23)		0	gi|4557727|ref|NP_000228.1|	lipoprotein lipase precursor
HG100032 3N 0_160000_ge ne_prediction 2	1347	no_pf am		1 (1-23)		0	gi|1619760 0|gb|AAL13166.1|	type V preprocollagen alpha 2 chain
HG100032 7N 0_1000_gene _prediction1	527	SAP A SapB_ 1 SapB _2		1 (1-17)		0	gi|27676136|ref|XP_223521.1|	similar to prosaposin (variant Gaucher

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000327N 0_160000_ge ne_prediction 1	527	SAP A SapB _1 SapB _2	1	1 (1-17)		0	gi|27676136|ref|XP_223521.1|	similar to prosaposin (variant Gaucher
HG1000434N 0_160000_ge ne_prediction 1		no_pf am	1	1 (1-19)		0		no_human_hit
HG1000449N 0_160000_ge ne_prediction 1	84	trefoil	1	1 (1-21)		0	gi|4507451|ref|NP_003216.1|	trefoil factor 1 (breast cancer,
HG1000807N 0_160000_ge ne_prediction 1	278	kazal ig	1	1 (1-24)		0	gi|17384405|emb|CAD13245.1| bA113O24.1	(similar to insulin-like growth factor binding protein)
HG1000807N 0_5000_gene prediction1	278	kazal	1	1 (1-24)		0	gi|17384405|emb|CAD13245.1| bA113O24.1	(similar to insulin-like growth factor binding

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
_prediction1								protein)
HG1001280N0_160000_gene_prediction1	1193	no_pfam	1	(1-22)		0	gi|665825|dbj|BAA13448.1|	Similar to Human C219-reactive peptide
HG1000193N0_160000_gene_prediction1	879	no_pfam	0.99	(1-16)		0	gi|24797158|ref|NP_005899.2|	mannosidase, beta A, lysosomal;
HG1000286N0_160000_gene_prediction1	418	serpin	0.99	(1-21)		0	gi|20141241|sp|P50454|CBP2_HUMAN	Collagen-binding protein 2
HG1000992N0_160000_gene_prediction1	645	no_pfam	0.99	(1-22)		0	gi|27477294|ref|XP_209510.1|	similar to solute carrier family 9,
HG1001148N0_160000_gene_prediction1	814	Pep_M12B_propep Reprolysin	0.99	(1-17)		0	gi|15778976|gb|AAH14566.1|AAH14566	a disintegrin and metalloproteinase domain 15 (metargidin)
HG1001185N0_160000_gene_prediction	502	Neur_chan_LBD	0.99	(1-22)		0	gi|4502831|ref|NP_000737.1|	cholinergic receptor, nicotinic,

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
2								
HG1001280N 0_5000_gene prediction1	110	no_pf am	0.99	(1-22)		0	gi|20964036|ref|XP_135708.1|	similar to melanoma inhibitory
HG1001302N 0_160000_ge ne_prediction 2	956	EGF vwa	0.99	(1-23)		0	gi|14549163|sp|O00339|MTN2_ HUMAN	Matrilin-2 precursor
HG1000361N 0_160000_ge ne_prediction 1	697	rvt LRR CT rnase H ig	0.98	(1-24)		0	gi|24233577|ref|NP_700356.1|	hypothetical protein FLJ90440
HG1000361N 0_20000_gen e_prediction1	1,065	no_pf am	0.98	(1-24)		0	gi|7662320|ref|NP_055628.1|	KIAA0806 gene product [Homo sapiens]
HG1000792N 0_160000_ge ne_prediction 1	194	no_pf am	0.98	(1-18)		0	gi|22749301|ref|NP_689850.1|	hypothetical protein MGC17301
HG1000934N 0_160000_ge ne_prediction 1	697	rvt LRR CT rnase H ig	0.98	(1-24)		0	gi|24233577|ref|NP_700356.1|	hypothetical protein FLJ90440

289

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000976N 0_160000_gene_prediction 1	524	p450	0.98	(8-33)		0	gi|20532036|sp|Q9HCS2|CPFC_HUMAN	Cytochrome P450 4F12 (CYPIVF12)
HG1000992N 0_10000_gene_prediction1	645	no_pfam	0.98	(1-22)		0	gi|27477294|ref|XP_209510.1|	similar to solute carrier family 9,
HG1001185N 0_1000_gene_prediction1	502	Neur_chan_LBD	0.98	(1-22)		0	gi|4502831|ref|NP_000737.1|	cholinergic receptor, nicotinic,
HG1001185N 0_160000_gene_prediction 1	502	Neur_chan_LBD	0.98	(1-22)		0	gi|4502831|ref|NP_000737.1|	cholinergic receptor, nicotinic,
HG1001185N 0_1000_gene_prediction2	502	Neur_chan_LBD	0.98	(1-22)		0	gi|4502831|ref|NP_000737.1|	cholinergic receptor, nicotinic,
HG1001185N 0_5000_gene_prediction1	502	Neur_chan_LBD	0.98	(1-22)		0	gi|4502831|ref|NP_000737.1|	cholinergic receptor, nicotinic,
HG1001280N 0_10000_gene_prediction1	1415	no_pfam	0.98	(1-22)		0	gi|27477706|ref|XP_046126.4|	similar to Meningioma-expressed antigen
HG1000361N 0_10000_gene_prediction1	1094	no_pfam	0.97	(1-24)		0	gi|24308087|ref|NP_056356.1|	leucine-rich repeats and
HG1001381N 0_1000_gene	1110	no_pfam	0.96	(1-22)		0	gi|6090615|gb|AAF03259.1|	dihydropyridine receptor alpha 2

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
_prediction1								
HG1000263N0_5000_gene _prediction1	112	no_pf am	0.95	(1-19)		0	gi|7661696|ref|NP_054796.1|	DKFZP586O0120 protein [Homo sapiens]
HG1000579N0_160000_ge ne_prediction 1	347	Glyco _trans _f_6	0.89	(1-18)		0	gi|21619110|gb|AAH32499.1|	Similar to Forssman glycolipid
HG1000191N0_160000_ge ne_prediction 1	116	no_pf am	0.79	(1-18)		0	gi|27479383|ref|XP_208098.1|	similar to RE52392p [Drosophila
HG1000296N0_160000_ge ne_prediction 2	912	no_pf am	0.77	(2-23)		0	gi|3327036|dbj|BAA31586.1|	KIAA0611 protein [Homo sapiens]
HG1000346N0_1000_gene _prediction1	354	no_pf am	0.75	(14-37)		0	gi|27754174|ref|NP_776169.1|	hypothetical protein MGC46680 [Homo
HG1000963N0_5000_gene _prediction1	234	no_pf am	0.74	(1-20)		0	gi|14749486|ref|XP_051854.1|	similar to Mesoderm development
HG1000610N0_160000_ge ne_prediction 1	131	no_pf am	0.7	(1-14)		0	gi|17437412|ref|XP_065554.1|	hypothetical protein XP_065554 [Homo

291

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000342N 0_160000_ge ne_prediction 1	80	no_pf am	0.69	(1-26)	(47-66)	1	gi|7019343|ref|NP_037362.1|	CD24 antigen (small cell lung
HG1000342N 0_160000_ge ne_prediction 2	80	no_pf am	0.69	(1-26)	(47-66)	1	gi|7019343|ref|NP_037362.1|	CD24 antigen (small cell lung
HG1000650N 0_20000_gen e_prediction1		no_pf am	0.68	(1-19)		0		no_human_hit
HG1000191N 0_160000_ge ne_prediction 2	116	no_pf am	0.67	(1-18)		0	gi|27479383|ref|XP_208098.1|	similar to RE52392p [Drosophila
HG1000449N 0_160000_ge ne_prediction 3	80	trefoil	0.67	(15-37)		0	gi|585328|sp|Q07654|TFF3_HU MAN	Trefoil factor 3 precursor
HG1000181N 0_20000_gen e_prediction1	1342	zf-C2H2	0.66	(1-23)		0	gi|11560152|ref|NP_071378.1|	zinc finger protein 335;
HG1001058N 0_160000_ge ne_prediction 1	748	no_pf am	0.65	(1-23)		0	gi|28829101|gb|AAM34346.2|	similar to Homo sapiens (Human). Ankyrin
HG1000187N 0_160000_ge		no_pf am	0.63	(2-34)		0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
0_160000_ge ne_prediction 2		am						
HG1000191N 0_1000_gene _prediction1	116	no_pf am	0.62	(1-18)		0	gi|27479383|ref|XP_208098.1|	similar to RE52392p [Drosophila
HG1000319N 0_160000_ge ne_prediction 1		no_pf am	0.61	(1-16)		0		no_human_hit
HG1000137N 0_0_gene_pre diction1	163	no_pf am	0.58	(1-19)	(25-47)	1	gi|13124770|ref|NP_076869.1|	hypothetical protein IMAGE3455200
HG1000191N 0_5000_gene prediction1	116	no_pf am	0.57	(1-18)		0	gi|27479383|ref|XP_208098.1|	similar to RE52392p [Drosophila
HG1001350N 0_5000_gene prediction1	4957	no_pf am	0.51	(13-30)		0	gi|7512280|pir|T03455	ALR protein - human

293

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000327N 0_160000_ge ne_prediction 2	527	SAP_A SapB_1 SapB_2	0.48	(1-25)		0	gi|2676136|ref|XP_223521.1|	similar to prosaposin (variant Gaucher
HG1000179N 0_160000_ge ne_prediction 1	143	no_pf am	0.46	(1-25)		0	gi|13436134|gb|AAH04884.1|A AH04884	Unknown (protein for MGC:11141)
HG1000991N 0_160000_ge ne_prediction 1	381	PA	0.43	(16-41)		0	gi|6005864|ref|NP_009213.1|	ring finger protein 13; RING zinc

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001038N 0_5000_gene _prediction1	433	ALG 3	0.4	(1-22)	(80-102)(138-160)	2	gi|13279206|gb|AAH04313.1|A AH04313	Unknown (protein for
HG1001376N 0_160000_ge ne_prediction 2	1353	no_pf am	0.39	(1-15)		0	gi|7512821|pir|T00347	hypothetical protein DKFZp566G1246.1,
HG1001376N 0_20000_gen	1320	no_pf am	0.38	(1-15)		0	gi|22052701|ref|XP_051956.8|	similar to hypothetical protein

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
e_prediction2								
HG1000409N 0_160000_ge ne_prediction 1	1259	rvt	0.35	(12-28)		0	gi|126295|sp|P08547|LIN1_HU MAN	LINE-1 REVERSE TRANSCRIPTASE HOMOLOG
HG1000884N 0_160000_ge ne_prediction 1	716	no_pf am	0.35	(1-25)	(120-142)(231-253)(255-272)	3	gi|28478194|ref|XP_133389.3|	similar to hypothetical protein
HG1000575N 0_160000_ge ne_prediction 1		no_pf am	0.34	(11-41)		0		no_human_hit
HG1000906N 0_10000_gen e_prediction1	1211	Arma dillo_ seg	0.32	(3-18)		0	gi|4505843|ref|NP_003619.1|	plakophilin 4 [Homo sapiens]
HG1000485N 0_160000_ge ne_prediction 1	312	CBM _21	0.29	(10-33)		0	gi|4927640|gb|AAD33215.1|	PPP1R5 [Homo sapiens]
HG1000328N 0_160000_ge ne_prediction 1	1299	no_pf am	0.28	(16-40)		0	gi|7662198|ref|NP_055647.1|	TBC1 domain family, member 4; KIAA0603

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000231N0_160000_gene_prediction1	214	abhydrolase_2	0.27	(1-18)		0	gi|4679012|gb|AAD26994.1|	lysophospholipase isoform [Homo sapiens]
HG1001257N0_10000_gene_prediction1	811	no_pfam	0.27	(1-17)	(39-61)(71-93)(198-220)	3	gi|18676478|dbj|BAB84891.1|	FLJ00136 protein [Homo sapiens]
HG1000026N0_5000_gene_prediction1	737	ABC_membrane ABC_tran	0.24	(1-26)		0	gi|12248755|dbj|BAB20265.1|	mono ATP-binding cassette protein [Homo
HG1000300N0_160000_gene_prediction1	174	pro_isomerase	0.23	(1-15)		0	gi|12804335|gb|AAH03026.1|AAH03026	Unknown (protein for
HG1000109N0_160000_gene_prediction1	757	pkinase_C pkinase	0.22	(1-25)		0	gi|28570186|ref|NP_060618.2|	WINS1 protein [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001110N0_160000_gene_prediction1	757	pkinase_C pkinase	0.22	(1-25)		0	gi|28570186|ref|NP_060618.2|	WINS1 protein [Homo sapiens]
HG1001376N0_160000_gene_prediction3	1353	no_pfam	0.22	(1-37)		0	gi|7512821|pir||T00347	hypothetical protein DKFZp566G1246.1,
HG1000026N0_20000_gene_prediction1	737	ABC_membrane ABC_tran	0.21	(1-26)		0	gi|12248755|dbj|BAB20265.1|	mono ATP-binding cassette protein [Homo
HG1000276N0_1000_gene_prediction1	108	no_pfam	0.21	(1-25)		0	gi|8923930|ref|NP_060934.1|	uncharacterized hematopoietic
HG1000822N0_160000_gene_prediction2	482	Hist_deacetyl	0.2	(18-36)		0	gi|13128860|ref|NP_004955.2|	histone deacetylase 1; reduced
HG1000173N0_20000_gene_prediction1	259	no_pfam	0.19	(1-20)		0	gi|21751939|dbj|BAC04076.1|	unnamed protein product [Homo sapiens]
HG1001044N0_1000_gene_prediction1	1121	no_pfam	0.19	(1-31)		0	gi|17975763|ref|NP_004526.1|	myelin transcription factor 1;
HG1000299N0_1000_gene	459	no_pfam	0.18	(19-34)		0	gi|4503729|ref|NP_002005.1|	FK506-binding protein 4;

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
0_1000_gene_prediction1		am						FK506-binding
HG1000659N0_160000_gene_prediction_1	550	no_pfam	0.16	(5-28)	(85-107)(117-139)(159-181)(196-218)(230-252)	5	gi|27498395|ref|XP_210662.1|	similar to hypothetical protein;
HG1000659N0_160000_gene_prediction_2	550	no_pfam	0.16	(5-28)	(85-107)(117-139)(159-181)(196-218)(230-252)	5	gi|27498395|ref|XP_210662.1|	similar to hypothetical protein;
HG1000013N0_160000_gene_prediction_1		no_pfam	0.16	(13-30)		0		no_human_hit
HG1000173N0_160000_gene_prediction_1	259	no_pfam	0.16	(1-20)		0	gi|21751939|dbj|BAC04076.1|	unnamed protein product [Homo sapiens]
HG1000330N0_160000_gene_prediction_1	765	no_pfam	0.16	(13-27)		0	gi|11225260|ref|NP_003277.1|	DNA topoisomerase I; type I DNA
HG1000178N0_10000_gen	188	pfkB	0.15	(1-32)		0	gi|22035592|ref|NP_057706.2|	mitochondrial ribosomal protein L35

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
e_prediction1								
HG1000178N0_10000_gen e_prediction2	188	pfkB	0.15	(1-32)		0	gi|22035592|ref|NP_057706.2|	mitochondrial ribosomal protein L35
HG1000640N0_160000_ge ne_prediction1	194	no_pf am	0.15	(1-22)		0	gi|15011990|gb|AAH10889.1|A AH10889	Similar to CGI-116 protein [Homo
HG1001000N0_160000_ge ne_prediction1	167	no_pf am	0.15	(1-25)	(124-146)	1	gi|27483462|ref|XP_208106.1|	similar to hypothetical protein
HG1001418N0_160000_ge ne_prediction1	508	no_pf am	0.15	(19-33)		0	gi|27698034|ref|XP_223654.1|	similar to HSPC159 protein [Homo
HG1000153N0_20000_gen e_prediction1	960	kinesi n 60s_ri boso mal	0.14	(16-40)		0	gi|20143967|ref|NP_612565.1|	kinesin-like 5 isoform 1; mitotic
HG1000255N0_160000_ge ne_prediction1	122	no_pf am	0.13	(1-17)		0	gi|3212110|emb|CAA76759.1|	prefoldin subunit 1 [Homo sapiens]
HG1000186N0_160000_ge	452	no_pf am	0.12	(1-19)	(34-56)	1	gi|10092615|ref|NP_061108.2|	ethanolamine kinase [Homo sapiens]

300

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
ne_prediction 1								
HG1000259N 0_160000_ge ne_prediction 1	112	no_pf am	0.12	(1-19)		0	gi|766169|ref|NP_054796.1|	DKFZP586O0120 protein [Homo sapiens]
HG1000084N 0_10000_gen e_prediction1	879	no_pf am	0.11	(1-19)		0	gi|21264318|ref|NP_004028.3|	adenosine monophosphate deaminase 2
HG1000217N 0_160000_ge ne_prediction 1	99	no_pf am	0.11	(14-37)		0	gi|18577116|ref|XP_084521.1|	hypothetical protein XP_084521 [Homo
HG1000217N 0_160000_ge ne_prediction 2	99	no_pf am	0.11	(14-37)		0	gi|18577116|ref|XP_084521.1|	hypothetical protein XP_084521 [Homo
HG1000329N 0_160000_ge ne_prediction 1	312	no_pf am	0.11	(17-33)		0	gi|5668598|gb|AAD45972.1|AF 106062_1	Wiskott-Aldrich syndrome
HG1000227N 0_160000_ge ne_prediction 1	169	Sdh_c yt	0.1	(1-23)		0	gi|4506863|ref|NP_002992.1|	succinate dehydrogenase complex,
HG1000269N 0_10000_gen	121	Yippe e	0.1	(1-15)		0	gi|7706341|ref|NP_057145.1|	yippee protein [Homo sapiens]

301

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
e_prediction1								
HG1000990N0_160000_gene_prediction1	234	no_pfam	0.1	(1-20)	(203-225)	1	gi|8924262|ref|NP_061113.1|	triggering receptor expressed on
HG1000998N0_160000_gene_prediction1		no_pfam	0.1	(6-24)		0		no_human_hit
HG1001225N0_160000_gene_prediction1		no_pfam	0.1	(1-17)		0		no_human_hit
HG1001269N0_5000_gene_prediction1	256	ribonuclease_T2	0.1	(16-30)		0	gi|5231228|ref|NP_003721.2|	ribonuclease 6 precursor [Homo sapiens]
HG1001269N0_160000_gene_prediction1	256	ribonuclease_T2	0.1	(16-30)		0	gi|5231228|ref|NP_003721.2|	ribonuclease 6 precursor [Homo sapiens]
HG1000103N0_160000_gene_prediction1	803	HSP90	0.09	(8-22)		0	gi|4507677|ref|NP_003290.1|	tumor rejection antigen (gp96) 1; Tumor
HG1000143N0_1000_gene_prediction1	194	Ribosomal_S4	0.09	(1-19)		0	gi|14141193|ref|NP_001004.2|	ribosomal protein S9; 40S ribosomal

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000396N0_160000_gene_prediction1	1275	no_pfam	0.09	(18-31)		0	gi|5052951|gb|AAD38785.1|AF149422_2	unknown [Homo sapiens]
HG1000066N0_160000_gene_prediction1	216	ras	0.08	(1-17)		0	gi|28376635|ref|NP_783865.1|	Rab37-like [Homo sapiens]
HG1000078N0_1000_gene_prediction1	122	no_pfam	0.08	(1-25)		0	gi|27663408|ref|XP_216594.1|	similar to cell division cycle 2-like
HG1000117N0_160000_gene_prediction1	420	PX	0.08	(1-12)		0	gi|7512733|pir||T08691	hypothetical protein DKFZp564F052.1 - human
HG1000157N0_160000_gene_prediction1	255	14-3-3	0.08	(19-32)		0	gi|5803225|ref|NP_006752.1|	tyrosine 3/tryptophan 5 - monooxygenase
HG1000194N0_160000_gene_prediction1	99	UPF0136	0.08	(14-42)	(22-41)	1	gi|7662639|ref|NP_054770.1|	PTD011 protein [Homo sapiens]
HG1000228N0_40000_gene_prediction1	108	no_pfam	0.07	(1-25)		0	gi|2160382|dbj|BAA04130.1|	dihydrolipoamide succinyltransferase
HG1000228N0_20000_gen	108	no_pfam	0.07	(1-25)		0	gi|2160382|dbj|BAA04130.1|	dihydrolipoamide succinyltransferase

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
e_prediction1								
HG1000228N 0_160000_gene_prediction 1	108	no_pf am	0.07	(1-25)		0	gi|2160382|dbj|BAA04130.1|	dihydrolipoamide succinyltransferase
HG1000409N 0_10000_gene_prediction1		no_pf am	0.07	(18-41)		0		no_human_hit
HG1000611N 0_160000_gene_prediction 1	3113	no_pf am	0.07	(14-36)		0	gi|14670381|ref|NP_057427.2|	centromere protein F (350/400kD);
HG1000015N 0_0_gene_pre diction1	2322	lamin.in_G	0.06	(14-31)		0	gi|4503099|ref|NP_001888.1|	melanoma-associated chondroitin sulfate
HG1000088N 0_5000_gene_prediction1	531	PK	0.06	(1-23)		0	gi|478822|pir||S30038	pyruvate kinase (EC 2.7.1.40), muscle splice
HG1000143N 0_10000_gene_prediction1	277	adh_s hort Ribos omal_S4	0.06	(1-19)		0	gi|4502601|ref|NP_001227.1|	carbonyl reductase 3; carbonyl

304

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000167N 0_5000_gene _prediction1	154	no_pf am	0.06	(16-29)		0	gi|16877034|gb|AAH16790.1|A AH16790	Unknown (protein for MGC:24015)
HG1000243N 0_5000_gene _prediction1	90	HMG 14_17	0.06	(16-32)		0	gi|18555712|ref|XP_096198.1|	hypothetical protein XP_096198 [Homo
HG1000825N 0_160000_ge ne_prediction 1	378	no_pf am	0.06	(1-30)		0	gi|27703812|ref|XP_225057.1|	similar to transcription elongation
HG1001019N 0_1000_gene _prediction1	197	DUF7 98	0.06	(19-36)		0	gi|27479097|ref|XP_209931.1|	similar to hypothetical protein,
HG1000044N 0_160000_ge ne_prediction 1	1260	DIL	0.05	(1-26)		0	gi|13431718|sp|Q9ULV0|MY5B _HUMAN	Myosin Vb (Myosin 5B)
HG1000100N 0_10000_gen e_prediction1	318	Pribo syltra n	0.05	(18-34)		0	gi|4506127|ref|NP_002755.1|	phosphoribosyl pyrophosphate synthetase

305

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000149N 0_160000_ge ne_prediction 1		no_pf am	0.05	(15-44)		0		no_human_hit
HG1000183N 0_1000_gene prediction1	553	no_pf am	0.05	(16-36)		0	gi|18148873|dbj|BAB83517.1|	hUST3 [Homo sapiens]
HG1000183N 0_160000_ge ne_prediction 2	553	no_pf am	0.05	(16-36)		0	gi|18148873|dbj|BAB83517.1|	hUST3 [Homo sapiens]
HG1000213N 0_5000_gene prediction1	89	BAF	0.05	(1-17)		0	gi|4502389|ref|NP_003851.1|	barrier to autointegration factor;
HG1000294N 0_5000_gene prediction1	126	no_pf am	0.05	(18-36)		0	gi|11386175|ref|NP_068778.1|	protein phosphatase 1, regulatory
HG1000430N 0_160000_ge ne_prediction 1	158	no_pf am	0.05			0	gi|14754198|ref|XP_045450.1|	similar to hypothetical protein (L1H 3
HG1000078N 0_5000_gene prediction1	122	no_pf am	0.04	(1-25)		0	gi|27663408|ref|XP_216594.1|	similar to cell division cycle 2-like
HG1000139N 0_5000_gene	144	no_pf am	0.04			0	gi|13375725|ref|NP_078834.1|	chromosome 14 open

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
0_5000_gene_prediction1		am						reading frame 138
HG1000143N0_160000_ge ne_prediction 1	277	adh_s hort Ribos omal_ S4	0.04	(1-19)		0	gi|4502601|ref|NP_001227.1|	carbonyl reductase 3; carbonyl
HG1000162N0_160000_ge ne_prediction 1	211	Ribos omal_ L13e	0.04	(1-34)		0	gi|15431295|ref|NP_150254.1|	ribosomal protein L13; 60S ribosomal
HG1000168N0_160000_ge ne_prediction 1	213	no_pf am	0.04	(13-28)		0	gi|1710488|sp|P50914|RL14_H UMAN	60S ribosomal protein L14 (CAG-ISL
HG1000187N0_160000_ge ne_prediction 1	338	Trans posas e_22	0.04			0	gi|2072963|gb|AAC51270.1|	p40 [Homo sapiens]
HG1000247N0_160000_ge ne_prediction 1	843	RGS	0.04	(1-24)		0	gi|4757824|ref|NP_004646.1|	axin 2; axil [Homo sapiens]
HG1000273N0_160000_ge ne_prediction 2	151	no_pf am	0.04			0	gi|10438240|dbj|BAB15204.1|	unnamed protein product [Homo sapiens]

307

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000539N 0_160000_ge ne_prediction 1	850	no_pf am	0.04			0	gi|25051344|ref|XP_146880.2|	similar to hypothetical protein
HG1000539N 0_160000_ge ne_prediction 2	850	no_pf am	0.04			0	gi|25051344|ref|XP_146880.2|	similar to hypothetical protein
HG1000560N 0_160000_ge ne_prediction 1		no_pf am	0.04		(75-97)	1		no_human_hit
HG1000740N 0_160000_ge ne_prediction 1	719	no_pf am	0.04			0	gi|20864261|ref|XP_143091.1|	similar to hypothetical protein
HG1000020N 0_5000_gene _prediction1	121	no_pf am	0.03			0	gi|7263959|emb|CAB81642.1|	bA395L14.5 (novel phosphoglucomutase
HG1000084N 0_5000_gene _prediction1	879	no_pf am	0.03	(1-33)		0	gi|21264318|ref|NP_004028.3|	adenosine monophosphate deaminase 2
HG1000135N 0_5000_gene _prediction1	130	HesB -like	0.03	(18-33)		0	gi|13569911|ref|NP_112202.1|	hypothetical protein MGC4276 similar
HG1000169N 0_20000_gen e_prediction1	370	amino tran_ 5	0.03	(1-23)		0	gi|17402893|ref|NP_478059.1|	phosphoserine aminotransferase,

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000169N0_160000_gene_prediction_1	370	amino tran_5	0.03	(1-23)		0	gi|17402893|ref|NP_478059.1|	phosphoserine aminotransferase,
HG1000189N0_160000_gene_prediction_1	145	no_pfam	0.03	(4-19)		0	gi|27485699|ref|XP_209373.1|	similar to agCP11542 [Anopheles
HG1000189N0_160000_gene_prediction_2	145	no_pfam	0.03	(4-19)		0	gi|27485699|ref|XP_209373.1|	similar to agCP11542 [Anopheles
HG1000246N0_5000_gene_prediction1	608	no_pfam	0.03			0	gi|5630076|gb|AAD45821.1|AC006017_1	
HG1000248N0_0_gene_prediction1	165	cofilin_ADF	0.03	(12-34)		0	gi|5802966|ref|NP_006861.1|	destrin (actin depolymerizing factor);
HG1000288N0_10000_gene_prediction1	343	Transposase_1	0.03	(18-35)		0	gi|1263081|gb|AAC52010.1|	mariner transposase [Homo sapiens]
HG1000443N0_40000_gene_prediction1	1410	no_pfam	0.03	(1-15)		0	gi|11360251|pir||T47137	hypothetical protein DKFZp761K2213.1 - human
HG1000590N0_1000_gene_prediction1	167	no_pfam	0.03			0	gi|6841264|gb|AAF28985.1|AF161425_1	HSPC307 [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000626N0_160000_gene_prediction1		no_pfam	0.03	(14-43)		0		no_human_hit
HG1000871N0_160000_gene_prediction1	503	CH pkinase Activin_rec p	0.03		(364-386)	1	gi|3915750|sp|P37023|KIR3_HUMAN	Serine/threonine-protein kinase
HG1000959N0_10000_gene_prediction1	253	trans membrane4	0.03		(98-120)(212-234)	2	gi|21237748|ref|NP_004348.2|	CD151 antigen, isoform a;
HG1000961N0_160000_gene_prediction3	489	Peptidase_M16_C Peptidase_M16	0.03	(1-34)		0	gi|315348|gb|AAC15866.1|	mitochondrial processing peptidase beta

310

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000974N 0_5000_gene _prediction1	167	no_pf am	0.03		(29-51)	1	gi|6841264|gb|AAF28985.1|AF 161425_1	HSPC307 [Homo sapiens]
HG1001045N 0_160000_ge ne_prediction 1	338	Trans posas e_22	0.03			0	gi|5070621|gb|AAD39214.1|AF 148856_1	unknown [Homo sapiens]
HG1001110N 0_0_gene_pre diction1	277	pkina se	0.03	(1-17)		0	gi|10835065|ref|NP_002751.1|	protein kinase, Y-linked [Homo
HG1001223N 0_1000_gene _prediction1	299	sarco glyca n	0.03		(36-58)	1	gi|21040253|ref|NP_631906.1|	sarcoglycan zeta; zeta-sarcoglycan

311

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001281N0_160000_gene_prediction1	1797	no_pfam	0.03			0	gi|27481712|ref|XP_047961.5|	similar to HCH [Mus musculus] [Homo
HG1001317N0_5000_gene_prediction1	1380	ig	0.03	(1-26)		0	gi|10047201|dbj|BAB13394.1|	KIAA1568 protein [Homo sapiens]
HG1001017N0_10000_gene_prediction1		no_pfam	0.02	(2-12)		0		no_human_hit
HG1001017N0_1000_gene_prediction1		no_pfam	0.02	(2-12)		0		no_human_hit
HG1000014N0_160000_gene_prediction2	279	Na_K-ATPase	0.02		(100-117)	1	gi|4502281|ref|NP_001670.1|	ATPase, Na+/K+ transporting, beta 3
HG1000043N0_160000_gene_prediction3	477	7tm_1	0.02		(163-185)(197-219)(234-256)(302-324)	4	gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor
HG1000052N0_160000_gene_prediction1	1275	rvt Exo_endo_phos	0.02			0	gi|5052951|gb|AAD38785.1|AF149422_2	unknown [Homo sapiens]
HG1000084N0_5000_gene	268	no_pfam	0.02	(1-33)		0	gi|27478305|ref|XP_209869.1|	hypothetical protein XP_209869 [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
_prediction2								
HG1000093N 0_1000_gene prediction1	991	no_pf am	0.02	(1-9)		0	gi|8574029|emb|CAB94780.1| dJ1013A10.1	(PRP4 protein kinase
HG1000105N 0_160000_ge ne_prediction 1	398	cyclin _C	0.02			0	gi|4757930|ref|NP_004692.1|	cyclin B2 [Homo sapiens]
HG1000157N 0_1000_gene prediction1	255	14-3-3	0.02			0	gi|5803225|ref|NP_006752.1|	tyrosine 3/tryptophan 5 - monooxygenase
HG1000210N 0_40000_gen e_prediction1	535	PDZ	0.02			0	gi|21450785|ref|NP_004223.2|	suppressor of cytokine signaling 4;
HG1000242N 0_5000_gene prediction1	324	no_pf am	0.02	(1-11)		0	gi|3152378|emb|CAA73791.1|	DnaJ protein [Homo sapiens]
HG1000243N 0_5000_gene	90	HMG 14_17	0.02	(1-12)		0	gi|5031749|ref|NP_005508.1|	high-mobility group nucleosomal

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
_prediction2								
HG1000256N 0_160000_ge ne_prediction 1		no_pf am	0.02	(1-19)		0		no_human_hit
HG1000279N 0_0_gene_pre diction1	241	GST_ C	0.02	(8-28)		0	gi|14251209|ref|NP_001279.2|	chloride intracellular channel 1;
HG1000280N 0_5000_gene _prediction1	463	filam ent	0.02	(16-38)		0	gi|22065094|ref|XP_091665.3|	similar to RIKEN cDNA 4733401L19 [Mus
HG1000280N 0_5000_gene _prediction2	463	filam ent	0.02	(16-38)		0	gi|22065094|ref|XP_091665.3|	similar to RIKEN cDNA 4733401L19 [Mus
HG1000282N 0_160000_ge ne_prediction 1	142	no_pf am	0.02		(59-77)	1	gi|9910382|ref|NP_064628.1|	mitochondrial import receptor Tom22
HG1000292N 0_160000_ge ne_prediction 1	115	Ribos omal_ S26e	0.02			0	gi|13639392|ref|XP_017661.1|	similar to ribosomal protein S26 [Homo
HG1000313N 0_160000_ge ne_prediction 1	173	no_pf am	0.02			0	gi|4506283|ref|NP_003454.1|	protein tyrosine phosphatase type IVA,

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000330N0_20000_gene_prediction1	777	no_pfam	0.02			0	gi|4159885|gb|AAD05194.1|	unknown [Homo sapiens]
HG1000482N0_160000_gene_prediction1	207	no_pfam	0.02	(11-28)	(360-382)	1	gi|17484447|ref|XP_066102.1|	ribosomal protein L7a-like 2 [Homo
HG1000486N0_20000_gene_prediction1		no_pfam	0.02	(1-18)		0		no_human_hit
HG1000518N0_160000_gene_prediction1	226	no_pfam	0.02			0	gi|21756553|dbj|BAC04906.1|	unnamed protein product [Homo sapiens]
HG1000556N0_160000_gene_prediction1	473	no_pfam	0.02			0	gi|21750183|dbj|BAC03736.1|	unnamed protein product [Homo sapiens]
HG1000588N0_160000_gene_prediction1	977	no_pfam	0.02			0	gi|20178274|sp|O95782|A2A1_HUMAN	Adapter-related protein complex 2
HG1000600N0_160000_gene_prediction1		no_pfam	0.02			0		no_human_hit
HG1000648N0_160000_ge	828	no_pfam	0.02			0	gi|20521812|dbj|BAA86557.2|	KIAA1243 protein [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
ne_prediction 1								
HG1000696N 0_160000_ge ne_prediction 1	175	Trans posas e_22	0.02			0	gi|21758619|dbj|BAC05339.1|	unnamed protein product [Homo sapiens]
HG1000788N 0_160000_ge ne_prediction 1	820	LRR CT	0.02	(1-23)	(418-440)	1	gi|6168273|ref|XP_056282.1|	similar to KIAA1904 protein [Homo
HG1000874N 0_160000_ge ne_prediction 1	199	no_pf am	0.02			0	gi|13129102|ref|NP_077002.1|	hypothetical protein MGC955 [Homo
HG1000902N 0_20000_gen e_prediction1	531	cpn60_TCP _1	0.02	(1-17)		0	gi|14517632|dbj|BAB61032.1|	acute morphine dependence related
HG1000902N 0_160000_ge ne_prediction 2	531	cpn60_TCP _1	0.02	(1-17)		0	gi|14517632|dbj|BAB61032.1|	acute morphine dependence related
HG1000902N 0_1000_gene prediction1	531	cpn60_TCP _1	0.02	(1-17)		0	gi|14517632|dbj|BAB61032.1|	acute morphine dependence related
HG1000966N 0_1000_gene prediction1	350	no_pf am	0.02		(41-63)	1	gi|23270917|gb|AAH16849.1|	Similar to hypothetical protein MGC25511

316

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000966N0_5000_gene prediction1	350	no_pfam	0.02		(41-63)	1	gi|23270917|gb|AAH16849.1|	Similar to hypothetical protein MGC25511
HG1000994N0_160000_ge ne_prediction 1	869	7tm_1	0.02		(102-124)	1	gi|27498059|ref|XP_068166.3|	similar to olfactory receptor MOR145-2
HG1001014N0_160000_ge ne_prediction 3	477	7tm_1	0.02		(163-185)(197-219)(234-256)(302-324)	4	gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor
HG1001041N0_5000_gene prediction1	424	pkinase	0.02	(15-43)		0	gi|14776113|ref|XP_043047.1|	similar to Serine/threonine-protein
HG1001337N0_160000_ge ne_prediction 1	958	no_pfam	0.02			0	gi|7019521|ref|NP_037484.1|	squamous cell carcinoma antigen

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000151N0_160000_gene_prediction1	279	Pecanex_C	0.01		(39-61)(71-93)(137-159)(174-196)(217-239)(243-265)(296-318)(328-345)(382-404)(474-492)(505-527)(547-569)(576-598)	13	gi|22095363|ref|NP_071940.2|	chromosome 14 open reading frame 135
HG1000330N0_160000_gene_prediction3		no_pfam	0.01			0		no_human_hit
HG1000957N0_20000_gene_prediction1	1259	rvt	0.01		(657-676)(681-703)	2	gi|126295|sp|P08547|LIN1_HUMAN	LINE-1 REVERSE TRANSCRIPTASE HOMOLOG

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000960N 0_0_gene_pre diction1	628	SNF	0.01		(44-66)(122-139)(152-174)(201-223)(306-328)(416-438)(459-476)(491-513)(534-556)(583-605)	10	gi|16550619|dbj|BAB71018.1|	unnamed protein product [Homo sapiens]
HG1000960N 0_0_gene_pre diction2	628	SNF	0.01		(44-66)(122-139)(152-174)(201-223)(273-295)(383-405)(426-443)(458-480)(501-523)(550-572)	10	gi|16550619|dbj|BAB71018.1|	unnamed protein product [Homo sapiens]
HG1001280N 0_20000_gen e_prediction1	1415	no_pf am	0.01	(1-22)	(1438-1460)	1	gi|27477706|ref|XP_046126.4|	similar to Meningioma-expressed antigen
HG1000003N 0_10000_gen e_prediction1	1231	WW	0.01	(1-19)		0	gi|27708380|ref|XP_228858.1|	similar to KIAA1280 protein [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000041N 0_160000_ge ne_prediction 1	427	SPRY fn3	0.01	(15-45)		0	gi|27659604|ref|XP_226563.1|	similar to tripartite motif protein 9,
HG1000043N 0_160000_ge ne_prediction 2	477	7tm_ 1	0.01		(131-153)(165-187)(202-224)(270-292)	4	gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor
HG1000044N 0_5000_gene prediction1	1260	DIL	0.01			0	gi|13431718|sp|Q9ULV0|MY5B _HUMAN	Myosin Vb (Myosin 5B)
HG1000051N 0_160000_ge ne_prediction 1	467	IRF	0.01			0	gi|5453700|ref|NP_006138.1|	interferon regulatory factor 6 [Homo
HG1000057N 0_160000_ge ne_prediction 1	140	profili n	0.01			0	gi|4826898|ref|NP_005013.1|	profilin 1 [Homo sapiens]
HG1000060N 0_160000_ge ne_prediction 1	451	tubuli n_C tubuli n	0.01			0	gi|6755901|ref|NP_035783.1|	tubulin, alpha 1; tubulin alpha 1 [Mus
HG1000061N 0_10000_gen e_prediction1	179	arf	0.01	(16-32)		0	gi|6912244|ref|NP_036229.1|	ADP-ribosylation factor-5 [Homo

320

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000079N0_160000_gene_prediction1	227	ADK ADK_lid	0.01			0	gi|19923437|ref|NP_057366.2|	adenylate kinase 3 alpha like [Homo
HG1000098N0_160000_gene_prediction1	1061	guanylate_cyc ANF_receptor	0.01		(476-498)	1	gi|4505435|ref|NP_000897.1|	natriuretic peptide receptor
HG1000105N0_5000_gene_prediction1	398	cyclin_C	0.01			0	gi|4757930|ref|NP_004692.1|	cyclin B2 [Homo sapiens]
HG1000121N0_160000_gene_prediction1	228	no_pfam	0.01			0	gi|7706497|ref|NP_057392.1|	UMP-CMP kinase [Homo sapiens]
HG1000131N0_160000_gene_prediction1	239	no_pfam	0.01			0	gi|11545787|ref|NP_071356.1|	egl nine homolog 3; EGL nine
HG1000134N0_160000_gene_prediction1	268	no_pfam	0.01			0	gi|27479840|ref|XP_208277.1|	similar to heterogeneous nuclear

321

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000134N0_160000_gene_prediction2	268	no_pfam	0.01			0	gi|27479840|ref|XP_208277.1|	similar to heterogeneous nuclear
HG1000136N0_160000_gene_prediction1	327	Vps26	0.01	(11-28)		0	gi|3342000|gb|AAC39912.1|	H beta 58 homolog [Homo sapiens]
HG1000147N0_160000_gene_prediction1	204	Ribosomal_S7	0.01			0	gi|13904870|ref|NP_001000.2|	ribosomal protein S5; 40S ribosomal
HG1000166N0_160000_gene_prediction1	664	succ_DH_flav_C FAD_binding_2	0.01			0	gi|1169337|sp|P31040|DHSA_HUMAN	Succinate dehydrogenase
HG1000172N0_1000_gene_prediction1		cytochrome_c	0.01	(1-24)		0		no_human_hit
HG1000172N0_1000_gene_prediction2		cytochrome_c	0.01	(1-24)		0		no_human_hit
HG1000175N0_5000_gene_prediction1	977	no_pfam	0.01			0	gi|27710280|ref|XP_231752.1|	similar to hypothetical protein

322

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000175N0_10000_gene_prediction1	977	no_pfam	0.01			0	gi|27710280|ref|XP_231752.1|	similar to hypothetical protein
HG1000175N0_160000_gene_prediction1		no_pfam	0.01			0		no_human_hit
HG1000175N0_1000_gene_prediction1	977	no_pfam	0.01			0	gi|27710280|ref|XP_231752.1|	similar to hypothetical protein
HG1000192N0_160000_gene_prediction1	423	no_pfam	0.01	(17-41)		0	gi|20140802|sp|Q9GZL7|WD12_HUMAN	WD-repeat protein 12 (YTM1
HG1000193N0_160000_gene_prediction2	135	no_pfam	0.01		(152-174)	1	gi|27477958|ref|XP_209657.1|	similar to RH38554p [Drosophila
HG1000195N0_160000_gene_prediction1	128	no_pfam	0.01			0	gi|14602841|gb|AAH09922.1|AAH09922	Unknown (protein for
HG1000197N0_160000_gene_prediction1	1266	no_pfam	0.01		(140-162)	1	gi|20521055|dbj|BAA23714.2|	KIAA0442 [Homo sapiens]
HG1000202N0_20000_gen	1155	no_pfam	0.01			0	gi|28481020|ref|XP_129972.3|	similar to hypothetical protein [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
e_prediction1								
HG1000210N0_20000_gene_prediction1	535	PDZ	0.01			0	gi|21450785|ref|NP_004223.2|	suppressor of cytokine signaling 4;
HG1000218N0_1000_gene_prediction1	77	LIM	0.01	(15-30)		0	gi|4503047|ref|NP_001302.1|	cysteine-rich protein 1 (intestinal);
HG1000218N0_160000_gene_prediction1	77	LIM actin	0.01			0	gi|4503047|ref|NP_001302.1|	cysteine-rich protein 1 (intestinal);
HG1000218N0_10000_gene_prediction1	77	LIM	0.01	(16-30)		0	gi|4503047|ref|NP_001302.1|	cysteine-rich protein 1 (intestinal);
HG1000222N0_1000_gene_prediction1	98	no_pfam	0.01			0	gi|4505361|ref|NP_002482.1|	NADH dehydrogenase (ubiquinone) 1
HG1000233N0_1000_gene_prediction1	474	no_pfam	0.01			0	gi|4757762|ref|NP_004281.1|	ring finger protein 14; androgen
HG1000234N0_1000_gene_prediction1	474	IBR	0.01			0	gi|4757762|ref|NP_004281.1|	ring finger protein 14; androgen
HG1000234N0_160000_gene_prediction	474	IBR	0.01			0	gi|4757762|ref|NP_004281.1|	ring finger protein 14; androgen

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000238N 0_160000_ge ne_prediction 2	224	AhpC -TSA	0.01			0	gi|4758638|ref|NP_004896.1|	antioxidant protein 2; non-selenium
HG1000240N 0_160000_ge ne_prediction 1	391	NAP	0.01			0	gi|4758756|ref|NP_004528.1|	nucleosome assembly protein 1-like 1;
HG1000245N 0_160000_ge ne_prediction 1	339	UBX Trans posas e_22	0.01			0	gi|7022811|dbj|BAA91731.1|	unnamed protein product [Homo sapiens]
HG1000245N 0_5000_gene _prediction1	339	UBX	0.01	(18-39)		0	gi|7022811|dbj|BAA91731.1|	unnamed protein product [Homo sapiens]
HG1000249N 0_10000_gen e_prediction1	248	lectin _c	0.01			0	gi|14030460|gb|AAK52907.1|A F360991_1	mannan-binding lectin MBL
HG1000251N 0_160000_ge ne_prediction 1	195	efhan d	0.01	(17-30)		0	gi|6005731|ref|NP_009167.1|	calcium binding protein P22; SLC9A1
HG1000252N 0_5000_gene _prediction1	336	no_pf am	0.01			0	gi|6005747|ref|NP_009143.1|	ring finger protein 2 [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000254N 0_160000_ge ne_prediction 1	241	no_pf am	0.01	(1-26)		0	gi|7661790|ref|NP_054886.1|	HSPC128 protein [Homo sapiens]
HG1000262N 0_160000_ge ne_prediction 1	538	no_pf am	0.01			0	gi|20988566|gb|AAH30528.1|	Similar to KIAA1128 protein [Homo
HG1000264N 0_1000_gene prediction1	175	no_pf am	0.01	(1-32)		0	gi|7661786|ref|NP_054884.1|	HSPC125 protein [Homo sapiens]
HG1000264N 0_1000_gene _prediction2	175	no_pf am	0.01	(1-32)		0	gi|7661786|ref|NP_054884.1|	HSPC125 protein [Homo sapiens]
HG1000270N 0_20000_gen e_prediction1	222	SNF7	0.01			0	gi|7706353|ref|NP_057163.1|	CGI-149 protein [Homo sapiens]
HG1000270N 0_1000_gene _prediction1	222	SNF7	0.01			0	gi|7706353|ref|NP_057163.1|	CGI-149 protein [Homo sapiens]
HG1000274N 0_160000_ge ne_prediction 1	239	no_pf am	0.01		(59-76)(110-132)	0	gi|28193172|emb|CAD62328.1|	unnamed protein product [Homo sapiens]
HG1000276N 0_160000_ge ne_prediction 1	108	no_pf am	0.01			2	gi|8923930|ref|NP_060934.1|	uncharacterized hematopoietic

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000276N0_5000_gene_prediction1	108	no_pfam	0.01		(40-57)(91-113)	2	gi|8923930|ref|NP_060934.1|	uncharacterized hematopoietic
HG1000278N0_5000_gene_prediction1	160	no_pfam	0.01		(79-101)	1	gi|8924090|ref|NP_060979.1|	hypothetical protein PRO1855 [Homo
HG1000280N0_160000_gene_prediction_1	463	filament	0.01	(16-38)		0	gi|22065094|ref|XP_091665.3|	similar to RIKEN cDNA 4733401L19 [Mus
HG1000280N0_1000_gene_prediction1	463	filament	0.01	(16-38)		0	gi|22065094|ref|XP_091665.3|	similar to RIKEN cDNA 4733401L19 [Mus
HG1000280N0_160000_gene_prediction_2	463	filament	0.01	(16-38)		0	gi|22065094|ref|XP_091665.3|	similar to RIKEN cDNA 4733401L19 [Mus
HG1000280N0_1000_gene_prediction2	463	filament	0.01	(16-38)		0	gi|22065094|ref|XP_091665.3|	similar to RIKEN cDNA 4733401L19 [Mus
HG1000305N0_5000_gene_prediction1	396	Metallophos	0.01			0	gi|20070434|ref|NP_075563.2|	metallo phosphoesterase [Homo sapiens]
HG1000305N0_5000_gene_prediction2	396	Metallophos	0.01			0	gi|20070434|ref|NP_075563.2|	metallo phosphoesterase [Homo sapiens]

327

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000307N 0_160000_ge ne_prediction 1	219	NUDI X	0.01			0	gi|20455192|sp|Q9UKK9|NUD5 _HUMAN	ADP-sugar pyrophosphatase YSA1H
HG1000334N 0_160000_ge ne_prediction 1		no_pf am	0.01			0		no_human_hit
HG1000335N 0_160000_ge ne_prediction 1		no_pf am	0.01			0		no_human_hit
HG1000337N 0_5000_gene _prediction1	263	dUTP ase Ribos omal_ S4e	0.01			0	gi|4506725|ref|NP_000998.1|	ribosomal protein S4, X-linked X
HG1000372N 0_160000_ge ne_prediction 1	825	no_pf am	0.01			0	gi|4557669|ref|NP_000409.1|	interleukin 4 receptor precursor; CD124
HG1000397N 0_5000_gene _prediction1	242	no_pf am	0.01			0	gi|7705724|ref|NP_057041.1|	CGI-29 protein [Homo sapiens]
HG1000414N 0_160000_ge ne_prediction 1	1275	rvt Trans posas e_22	0.01			0	gi|2072977|gb|AAC51279.1|	putative p150 [Homo sapiens]

328

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000439N0_160000_gene_prediction1	263	Ribosomal_S4e	0.01	(1-18)		0	gi|4506725|ref|NP_000998.1|	ribosomal protein S4, X-linked X
HG1000449N0_20000_gene_prediction1	121	no_pfam	0.01	(1-14)		0	gi|27501128|ref|XP_210455.1|	similar to CG7874-PA [Drosophila
HG1000461N0_160000_gene_prediction1	1033	no_pfam	0.01			0	gi|21732487|emb|CAD38600.1|	hypothetical protein [Homo sapiens]
HG1000476N0_160000_gene_prediction1		no_pfam	0.01			0		no_human_hit
HG1000530N0_160000_gene_prediction1		no_pfam	0.01			0		no_human_hit
HG1000556N0_160000_gene_prediction2	473	no_pfam	0.01			0	gi|21750183|dbj|BAC03736.1|	unnamed protein product [Homo sapiens]
HG1000584N0_160000_gene_prediction1	971	no_pfam	0.01			0	gi|27688893|ref|XP_225599.1|	similar to hypothetical protein [Homo

329

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000587N 0_160000_ge ne_prediction 1	99	no_pf am	0.01			0	gi|18027742|gb|AAL55832.1|A F318325_1	unknown [Homo sapiens]
HG1000594N 0_160000_ge ne_prediction 1		no_pf am	0.01			0		no_human_hit
HG1000594N 0_160000_ge ne_prediction 2		no_pf am	0.01			0		no_human_hit
HG1000620N 0_160000_ge ne_prediction 1	362	no_pf am	0.01			0	gi|17455045|ref|XP_063008.1|	similar to KIAA1074 protein [Homo
HG1000631N 0_40000_gen e_prediction1	628	Furin-like Recep_L_d omain	0.01		(99-121)	0	gi|11494381|gb|AAG35790.1|A F288738_5	truncated epidermal growth
HG1000686N 0_160000_ge ne_prediction 1	338	Trans posas e_22	0.01			1	gi|5070621|gb|AAD39214.1|AF 148856_1	unknown [Homo sapiens]
HG1000712N 0_160000_ge ne_prediction	39	no_pf am	0.01			0	gi|7959817|gb|AAF71079.1|AF 116721_59	PRO1412 [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
1								
HG1000727N 0_160000_ge ne_prediction 1	73	no_pf am	0.01		(41-63)(126-145)	2	gi|6690244|gb|AAF24052.1|AF 090940_1	PRO0644 [Homo sapiens]
HG1000743N 0_160000_ge ne_prediction 2	510	KH	0.01			0	gi|4505425|ref|NP_002506.1|	neuro-oncological ventral antigen 1,
HG1000767N 0_5000_gene _prediction1	263	dUTP ase Ribos omal_ S4e	0.01			0	gi|4506725|ref|NP_000998.1|	ribosomal protein S4, X-linked X
HG1000822N 0_160000_ge ne_prediction 1	482	Hist_ deace tyl	0.01			0	gi|13128860|ref|NP_004955.2|	histone deacetylase 1; reduced
HG1000829N 0_160000_ge ne_prediction 1	755	WH2	0.01			0	gi|6539606|gb|AAF15947.1|	metastasis suppressor protein [Homo
HG1000860N 0_160000_ge ne_prediction 1	328	Pep_ M12 B_pro pep	0.01		(568-590)	1	gi|27484316|ref|XP_091055.3|	similar to disintegrin-like testicular

331

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
		disint egrin						
HG1000898N0_10000_gene prediction1	597	no_pfam	0.01			0	gi|27462211|gb|AAO15382.1|AF334386_1	multiple hat domains [Homo
HG1000898N0_160000_gene prediction1	597	no_pfam	0.01			0	gi|27462211|gb|AAO15382.1|AF334386_1	multiple hat domains [Homo
HG1000898N0_20000_gene prediction1	597	no_pfam	0.01			0	gi|27462211|gb|AAO15382.1|AF334386_1	multiple hat domains [Homo
HG1000902N0_160000_gene prediction1	531	cpn60_TCP1	0.01	(1-17)		0	gi|14517632|dbj|BAB61032.1|	acute morphine dependence related
HG1000906N0_20000_gene prediction1	1211	Armadillo_seg	0.01			0	gi|4505843|ref|NP_003619.1|	plakophilin 4 [Homo sapiens]
HG1000906N0_160000_gene prediction1	1211	Armadillo_seg	0.01			0	gi|4505843|ref|NP_003619.1|	plakophilin 4 [Homo sapiens]
HG1000921N0_5000_gene prediction1	730	SNF	0.01		(69-88)(98-120)(162-184)(253-275)(284...	5	gi|18202939|sp|Q9H2J7|NTT7_HUMAN	Orphan sodium- and

332

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
					275)(284-306)			
HG1000938N0_10000_gene_prediction1	421	no_pfam	0.01		(233-255)(270-292)(356-373)(378-400)(405-427)	5	gi|23138792|gb|AAH37878.1|	Unknown (protein for MGC:43868) [Homo
HG1000952N0_160000_gene_prediction1	1275	rvt	0.01			0	gi|2072953|gb|AAC51264.1|	putative p150 [Homo sapiens]
HG1000961N0_160000_gene_prediction1	108	rvt	0.01	(1-14)		0	gi|27478554|ref|XP_209815.1|	similar to Dumpy : shorter than
HG1000961N0_160000_gene_prediction2	591	no_pfam	0.01			0	gi|22209001|ref|NP_663748.1|	tigger transposable element derived 1;
HG1001000N0_160000_gene_prediction2	466	RCC1	0.01			0	gi|2981264|gb|AAC06338.1|	similar to golgi antigen; similar to
HG1001003N0_160000_ge	71	no_pfam	0.01			0	gi|2829147|gb|AAC00496.1|	lymphocyte-specific protein

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
0_160000_ge ne_prediction 1		am						1 [Homo
HG1001007N 0_160000_ge ne_prediction 1	1523	fn3	0.01			0	gi|35790|emb|CAA38068.1|	protein- [Homo
HG1001009N 0_0_gene_pre diction1	395	no_pf am	0.01			0	gi|17511700|gb|AAH18707.1|A AH18707	Unknown (protein for MGC:31765)
HG1001014N 0_160000_ge ne_prediction 2	477	7tm_ 1	0.01		(131- 153)(165- 187)(202- 224)(270- 292)	4	gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor
HG1001017N 0_40000_gen e_prediction1		no_pf am	0.01			0		no_human_hit
HG1001017N 0_20000_gen e_prediction1		no_pf am	0.01	(1-11)		0		no_human_hit
HG1001144N 0_160000_ge ne_prediction 1	338	Trans posas e_22	0.01			0	gi|5070621|gb|AAD39214.1|AF 148856_1	unknown [Homo sapiens]

334

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001172N 0_160000_ge ne_prediction 2	815	no_pf am	0.01			0	gi|21542539|gb|AAH33043.1|	echinoderm microtubule associated
HG1001253N 0_160000_ge ne_prediction 1	234	XRC C1_N	0.01			0	gi|27485468|ref|XP_170682.2|	similar to transient receptor
HG1001253N 0_160000_ge ne_prediction 2	234	XRC C1_N	0.01			0	gi|27485468|ref|XP_170682.2|	similar to transient receptor
HG1001267N 0_160000_ge ne_prediction 1	1275	Exo_ endo_ phos	0.01	(1-17)		0	gi|339777|gb|AAB59368.1|	ORF2 contains a reverse transcriptase
HG1001289N 0_160000_ge ne_prediction 1	1038	IBN_ NT	0.01			0	gi|5453998|ref|NP_006382.1|	importin 7; RAN-binding protein 7 [Homo
HG1001343N 0_10000_gen e_prediction1	1448	PI3_P I4_ki nase	0.01			0	gi|4758924|ref|NP_004561.1|	phosphoinositide-3-kinase, class 2,
HG1001343N 0_160000_ge ne_prediction 1	1448	PI3_P I4_ki nase	0.01			0	gi|4758924|ref|NP_004561.1|	phosphoinositide-3-kinase, class 2,

335

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001390N0_160000_gene_prediction_1	217	SH3	0.01			0	gi|4504111|ref|NP_002077.1|	growth factor receptor-bound protein
HG1001508N0_160000_gene_prediction_2	149	no_pfam	0.01			0	gi|18557931|ref|XP_087525.1|	similar to PRO1546 [Homo sapiens]
HG1000084N0_160000_gene_prediction_1		no_pfam	0			0		no_human_hit
HG1000084N0_160000_gene_prediction_2		no_pfam	0			0		no_human_hit
HG1000209N0_160000_gene_prediction_1	243	RWD	0		(49-71)(99-121)	2	gi|11065728|emb|CAC14427.1|	dJ493F7.2 (PTD013 similar to CGI-24
HG1000005N0_160000_gene_prediction_1	1275	rvt mito_carr	0			0	gi|5052951|gb|AAD38785.1|AF149422_2	unknown [Homo sapiens]
HG1000014N0_160000_gene_prediction_1	279	Na_K-ATPase	0		(33-55)	1	gi|4502281|ref|NP_001670.1|	ATPase, Na+/K+ transporting, beta 3

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000015N 0_160000_ge ne_prediction 1	445	GED	0			0	gi|15620899|dbj|BAB67813.1|	KIAA1920 protein [Homo sapiens]
HG1000015N 0_20000_gen e_prediction1	445	GED	0			0	gi|15620899|dbj|BAB67813.1|	KIAA1920 protein [Homo sapiens]
HG1000015N 0_5000_gene _prediction1	445	GED	0			0	gi|15620899|dbj|BAB67813.1|	KIAA1920 protein [Homo sapiens]
HG1000015N 0_160000_ge ne_prediction 2	445	GED lamin in_G	0			0	gi|15620899|dbj|BAB67813.1|	KIAA1920 protein [Homo sapiens]
HG1000020N 0_160000_ge ne_prediction 1	121	PGM_PM _M_I	0			0	gi|7263959|emb|CAB81642.1|	bA395L14.5 (novel phosphoglucomutase like protein)
HG1000020N 0_5000_gene prediction2	121	no_pf am	0			0	gi|7263959|emb|CAB81642.1|	bA395L14.5 (novel phosphoglucomutase like protein)
HG1000024N 0_10000_gen e_prediction1	273	no_pf am	0			0	gi|27486432|ref|XP_058770.2|	hypothetical protein XP_058770 [Homo
HG1000026N 0_160000_ge ne_prediction 1	737	ABC_mem brane ABC	0			0	gi|12248755|dbj|BAB20265.1|	mono ATP-binding cassette protein [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
		_tran						
HG1000030N 0_160000_gene_prediction 1	529	WD40	0			0	gi|16306496|ref|NP_387448.1|	F-box and WD-40 domain protein 1B,
HG1000039N 0_160000_gene_prediction 1	626	no_pfam	0	(16-39)		0	gi|20986052|ref|XP_142307.1|	similar to hypothetical protein
HG1000041N 0_5000_gene_prediction1	427	SPRY fn3	0			0	gi|27659604|ref|XP_226563.1|	similar to tripartite motif protein 9,
HG1000043N 0_160000_gene_prediction 1	477	7tm_1	0		(121-143)(155-177)(192-214)(260-282)	4	gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor
HG1000043N 0_5000_gene_prediction1	477	7tm_1	0		(121-143)(155-177)(192-214)(260-282)	4	gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor
HG1000044N 0_20000_gene_prediction1	1260	DIL	0			0	gi|13431718|sp|Q9ULV0|MY5B_HUMAN	Myosin Vb (Myosin 5B)

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000052N 0_160000_ge ne_prediction 2	454	no_pfam	0			0	gi|2226004|gb|AAB61714.1|	ORF1; MER37; putative transposase similar
HG1000052N 0_10000_gen e_prediction1	454	no_pfam	0			0	gi|2226004|gb|AAB61714.1|	ORF1; MER37; putative transposase similar
HG1000052N 0_20000_gen e_prediction1	454	no_pfam	0			0	gi|2226004|gb|AAB61714.1|	ORF1; MER37; putative transposase similar
HG1000058N 0_10000_gen e_prediction1	1259	ras	0	(8-24)		0	gi|126295|sp|P08547|LIN1_HU MAN	LINE-1 REVERSE TRANSCRIPTASE HOMOLOG
HG1000061N 0_5000_gene _prediction1	394	actin	0			0	gi|15778930|gb|AAH14546.1|A AH14546	Similar to ARP2 (actin-related
HG1000065N 0_5000_gene _prediction1	244	no_pfam	0			0	gi|13378141|ref|NP_054752.1|	DKFZP586A0522 protein [Homo sapiens]
HG1000065N 0_10000_gen e_prediction1	244	no_pfam	0			0	gi|13378141|ref|NP_054752.1|	DKFZP586A0522 protein [Homo sapiens]
HG1000065N 0_160000_ge ne_prediction 1	244	no_pfam	0			0	gi|13378141|ref|NP_054752.1|	DKFZP586A0522 protein [Homo sapiens]
HG1000068N 0_160000_ge		no_pfam	0			0		no_human_hit

339

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
ne_prediction 1								
HG1000070N 0_0_gene_pre diction1	405	pkina se	0		(62-84)	1	gi|18203885|gb|AAH21700.1|A AH21700	Unknown (protein for IMAGE:4052080)
HG1000073N 0_20000_gen e_prediction1	305	pkina se	0			0	gi|12803571|gb|AAH02618.1|A AH02618	serine/threonine kinase 16 [Homo
HG1000075N 0_160000_ge ne_prediction 1	1275	rvt	0			0	gi|2072972|gb|AAC51276.1|	putative p150 [Homo sapiens]
HG1000076N 0_160000_ge ne_prediction 1	149	efhan d	0			0	gi|4502549|ref|NP_001734.1|	calmodulin 2 (phosphorylase kinase,
HG1000081N 0_160000_ge ne_prediction 1	737	HSP9 0	0			0	gi|11277141|pir||T46243	hypothetical protein DKFZp761K0511.1 -
HG1000106N 0_160000_ge ne_prediction 1	367	GTP1 _OB G	0			0	gi|4758796|ref|NP_004138.1|	developmentally regulated GTP binding
HG1000107N 0_160000_ge ne_prediction 1	1939	Myos in_N myosi n hea	0	(1-16)		0	gi|476355|pir||A46762	myosin alpha heavy chain, cardiac muscle -

340

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
		d						
HG1000109N0_0_gene_prediction1	358	pkinase_C	0			0	gi|4826948|ref|NP_005035.1|	protein kinase, X-linked [Homo sapiens]
HG1000112N0_160000_gene_prediction1	1275	rvt actin	0			0	gi|2136112|pir||S65824	reverse transcriptase homolog - human
HG1000116N0_160000_gene_prediction1	737	HSP90 HATPase_c	0			0	gi|11277141|pir||T46243	hypothetical protein DKFZp761K0511.1 -
HG1000126N0_160000_gene_prediction1	88	Complex1_LYR	0			0	gi|10092689|ref|NP_065199.1|	hypothetical protein dJ122O8.2 [Homo
HG1000130N0_160000_gene_prediction1	209	HMG_box	0	(1-24)		0	gi|11321591|ref|NP_002120.1|	high-mobility group box 2;

341

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000132N0_160000_gene_prediction1	377	rrm	0		(33-55)(76-98)	2	gi|8671586|gb|AAF78291.1|AF107203_1	ataxin 2-binding protein [Homo
HG1000133N0_160000_gene_prediction1	119	no_pfam	0			0	gi|12654103|gb|AAH00864.1|AAH00864	Sjogren's syndrome nuclear
HG1000134N0_20000_gene_prediction1	268	no_pfam	0			0	gi|27479840|ref|XP_208277.1|	similar to heterogeneous nuclear
HG1000134N0_20000_gene_prediction2	268	no_pfam	0			0	gi|27479840|ref|XP_208277.1|	similar to heterogeneous nuclear
HG1000142N0_160000_gene_prediction1	294	Nucleoplasmin	0	(13-41)		0	gi|1831440408|gb|AAH21983.1|AAH21983	nucleophosmin (nucleolar
HG1000144N0_20000_gene_prediction1	137	Ribosomal_L28e	0			0	gi|13904866|ref|NP_000982.2|	ribosomal protein L28; 60S ribosomal

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000145N0_160000_gene_prediction1	137	Ribosomal_L28e	0	(1-25)		0	gi|13904866|ref|NP_000982.2|	ribosomal protein L28; 60S ribosomal
HG1000146N0_160000_gene_prediction1	204	Ribosomal_S7	0	0		0	gi|13904870|ref|NP_001000.2|	ribosomal protein S5; 40S ribosomal
HG1000150N0_10000_gene_prediction1	182	no_pfam	0	0		0	gi|14210488|ref|NP_115875.1|	dynactin 4 [Homo sapiens]
HG1000152N0_160000_gene_prediction1	345	no_pfam	0	0		0	gi|27481416|ref|XP_087026.2|	similar to GIG18 [Mus musculus] [Homo
HG1000161N0_160000_gene_prediction1	357	Transposase_22	0	0		0	gi|24308386|ref|NP_443169.1|	similar to hypothetical protein
HG1000163N0_160000_gene_prediction1	211	Ribosomal_L13e	0	0		0	gi|15431295|ref|NP_150254.1|	ribosomal protein L13; 60S ribosomal
HG1000164N0_5000_gene_prediction1	248	no_pfam	0	0		0	gi|7022560|dbj|BAA91644.1|	unnamed protein product [Homo sapiens]

343

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000165N 0_1000_gene prediction1	149	Stath min	0	(10-21)		0	gi|15680064|gb|AAH14353.1|A AH14353	Similar to stathmin
HG1000166N 0_160000_ge ne_prediction 2	1010	no_pf am	0	0		0	gi|14917097|sp|Q92622|Y226_ HUMAN	Hypothetical protein KIAA0226
HG1000167N 0_160000_ge ne_prediction 1	686	T-box	0	0		0	gi|22538470|ref|NP_005433.2|	eomesodermin; t box, brain, 2 [Homo
HG1000171N 0_40000_gen e_prediction1	235	no_pf am	0	0		0	gi|17448953|ref|XP_069792.1|	similar to LYRIC [Rattus norvegicus]
HG1000171N 0_160000_ge ne_prediction 1	235	no_pf am	0	0		0	gi|17448953|ref|XP_069792.1|	similar to LYRIC [Rattus norvegicus]
HG1000175N 0_160000_ge ne_prediction 2	384	no_pf am	0	0		0	gi|21753866|dbj|BAC04411.1|	unnamed protein product [Homo sapiens]
HG1000176N 0_1000_gene prediction1	818	zf-C2H2	0	0		0	gi|19718738|ref|NP_150630.1|	KRAB zinc finger protein KR18 [Homo
HG1000176N 0_160000_ge ne_prediction	818	zf-C2H2 KRA	0	0		0	gi|19718738|ref|NP_150630.1|	KRAB zinc finger protein KR18 [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
1		B						
HG1000177N0_160000_gene_prediction1	353	7tm_1	0		(35-57)(69-91)(106-128)(148-170)(190-212)(229-251)(266-288)	7	gi|4758240|ref|NP_004221.1|	endothelial differentiation,
HG1000178N0_160000_gene_prediction1	188	no_pfam	0			0	gi|22035592|ref|NP_057706.2|	mitochondrial ribosomal protein L35
HG1000178N0_160000_gene_prediction2	188	no_pfam	0			0	gi|22035592|ref|NP_057706.2|	mitochondrial ribosomal protein L35
HG1000180N0_1000_gene_prediction1	210	SAP	0			0	gi|18202440|sp|P82979|HCC1_HUMAN	Nuclear protein Hcc-1 (HSPC316)
HG1000181N0_10000_gene_prediction1	1342	zf-C2H2	0			0	gi|11560152|ref|NP_071378.1|	zinc finger protein 335;
HG1000181N0_160000_gene_prediction1	1342	zf-C2H2	0			0	gi|11560152|ref|NP_071378.1|	zinc finger protein 335;

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000183N0_160000_gene_prediction1	553	no_pfam	0			0	gi|18148873|dbj|BAB83517.1|	hUST3 [Homo sapiens]
HG1000186N0_20000_gene_prediction1		no_pfam	0	(1-7)		0		no_human_hit
HG1000186N0_160000_gene_prediction2	213	HMG_box	0	(1-25)		0	gi|27479871|ref|XP_208301.1|	similar to nonhistone chromosomal
HG1000187N0_20000_gene_prediction1		no_pfam	0	(1-7)		0		no_human_hit
HG1000187N0_160000_gene_prediction3	1275	no_pfam	0			0	gi|5052951|gb|AAD38785.1|AF149422_2	unknown [Homo sapiens]
HG1000189N0_1000_gene_prediction1	145	no_pfam	0			0	gi|27485699|ref|XP_209373.1|	similar to agCP11542 [Anopheles]
HG1000189N0_5000_gene_prediction1	145	no_pfam	0			0	gi|27485699|ref|XP_209373.1|	similar to agCP11542 [Anopheles]
HG1000189N0_1000_gene_prediction2	145	no_pfam	0			0	gi|27485699|ref|XP_209373.1|	similar to agCP11542 [Anopheles]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000189N 0_5000_gene _prediction2	145	no_pf am	0			0	gi|27485699|ref|XP_209373.1|	similar to agCP11542 [Anopheles
HG1000195N 0_10000_gen e_prediction1	128	no_pf am	0			0	gi|14602841|gb|AAH09922.1|A AH09922	Unknown (protein for
HG1000199N 0_160000_ge ne_prediction 1	763	3HC DH 3HC DH_ N	0			0	gi|20127408|ref|NP_000173.2|	hydroxyacyl-Coenzyme A
HG1000201N 0_10000_gen e_prediction1	193	no_pf am	0	(1-18)		0	gi|27477269|ref|XP_209223.1|	similar to Transforming protein RhoC
HG1000203N 0_5000_gene _prediction1	229	no_pf am	0			0	gi|23271386|gb|AAH36351.1|	Unknown (protein for MGC:35187) [Homo
HG1000204N 0_10000_gen e_prediction1	466	zf-C2H2	0			0	gi|24217433|gb|AAH38669.1|	Unknown (protein for MGC:41921) [Homo
HG1000209N 0_160000_ge ne_prediction 2	632	ank	0			0	gi|22060186|ref|XP_170722.1|	similar to KIAA0946 protein [Homo
HG1000215N 0_5000_gene _prediction1	198	no_pf am	0			0	gi|243521|dbj|BAA22536.1|	CIS2 [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000215N0_1000_gene prediction1	198	no_pf am	0			0	gi|2463521|dbj|BAA22536.1|	CIS2 [Homo sapiens]
HG1000219N0_10000_gen e prediction1	128	histon e	0			0	gi|27678050|ref|XP_225655.1|	similar to histone H2A.F/Z variant,
HG1000221N0_160000_ge ne prediction 1	5938	CH Plecti n	0			0	gi|15824727|gb|AAL09459.1|A F317696_1	macrophin 1 isoform 4 [Homo
HG1000221N0_20000_gen e prediction1	5938	CH Plecti n	0			0	gi|15824727|gb|AAL09459.1|A F317696_1	macrophin 1 isoform 4 [Homo
HG1000223N0_160000_ge ne prediction 1	100	Trans posas e_22	0			0	gi|20380930|gb|AAH28227.1|	Similar to hypothetical protein PRO2221
HG1000225N0_160000_ge ne prediction 1	239	PA28_beta	0	(1-18)		0	gi|28071124|emb|CAD61943.1|	unnamed protein product [Homo sapiens]
HG1000235N0_160000_ge ne prediction 1	118	V-ATPa se_G	0	(1-19)		0	gi|4757818|ref|NP_004879.1|	ATPase, H+ transporting, lysosomal,
HG1000236N0_160000_ge ne prediction	706	no_pf am	0			0	gi|25020161|ref|XP_207571.1|	similar to hypothetical protein

This page is not part of

the document!

US2003026780 / 2004-035732

4/4

Date: Apr 29, 2004

Recipient: IB

Address: Bureau international de l'OMPI 34, chemin des Colombettes Geneve 20 1211 CH

THIS PAGE BLANK (USPTO)

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000238N 0_160000_ge ne_prediction 1	224	AhpC -TSA	0	0 (13-33)		0	gi|4758638|ref|NP_004896.1|	antioxidant protein 2; non-selenium
HG1000238N 0_5000_gene prediction1	224	AhpC -TSA	0	0 (1-19)		0	gi|4758638|ref|NP_004896.1|	antioxidant protein 2; non-selenium
HG1000239N 0_160000_ge ne_prediction 1	391	NAP	0	0		0	gi|4758756|ref|NP_004528.1|	nucleosome assembly protein 1-like 1;
HG1000241N 0_160000_ge ne_prediction 1	118	no_pf am	0	0		0	gi|4759158|ref|NP_004588.1|	small nuclear ribonucleoprotein D2
HG1000243N 0_160000_ge ne_prediction 1	90	HMG 14_17	0	0 (17-37)		0	gi|5031749|ref|NP_005508.1|	high-mobility group nucleosomal
HG1000243N 0_160000_ge ne_prediction 2	90	HMG 14_17	0	0		0	gi|18555712|ref|XP_096198.1|	hypothetical protein XP_096198 [Homo
HG1000245N 0_1000_gene prediction1	339	UBX	0	0 (18-39)		0	gi|7022811|dbj|BAA91731.1|	unnamed protein product [Homo sapiens]

349

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000250N 0_160000_ge ne_prediction 1	763	3HC DH ECH 3HC DH_ N	0		(483-505)	1	gi|20127408|ref|NP_000173.2|	hydroxyacyl-Coenzyme A
HG1000252N 0_160000_ge ne_prediction 1	156	Ribos omal_ L23	0	(9-20)		0	gi|17105394|ref|NP_000975.2|	ribosomal protein L23a; 60S ribosomal
HG1000255N 0_10000_gen e_prediction1	122	KE2	0			0	gi|3212110|emb|CAA76759.1|	prefoldin subunit 1 [Homo sapiens]
HG1000262N 0_160000_ge ne_prediction 2	1275	rvt	0			0	gi|5070622|gb|AAD39215.1|AF 148856_2	unknown [Homo sapiens]
HG1000263N 0_160000_ge ne_prediction 1	112	no_pf am	0	(1-14)		0	gi|7661696|ref|NP_054796.1|	DKFZP586O0120 protein [Homo sapiens]
HG1000264N 0_5000_gene prediction1	175	no_pf am	0			0	gi|7661786|ref|NP_054884.1|	HSPC125 protein [Homo sapiens]
HG1000264N 0_5000_gene prediction2	175	no_pf am	0			0	gi|7661786|ref|NP_054884.1|	HSPC125 protein [Homo sapiens]

350

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000265N 0_160000_ge ne_prediction 1	168	ubiqu itin	0			0	gi|7661886|ref|NP_055579.1|	DAZ associated protein 2; KIAA0058 gene
HG1000266N 0_0_gene_pre diction1	145	no_pf am	0	(1-21)		0	gi|20559103|ref|XP_114160.1|	similar to [Segment 3 of 3] Lipin 3
HG1000266N 0_160000_ge ne_prediction 1	896	lipin_ N	0			0	gi|7662022|ref|NP_055461.1|	lipin 2 [Homo sapiens]
HG1000270N 0_160000_ge ne_prediction 1	222	SNF7	0			0	gi|7706353|ref|NP_057163.1|	CGI-149 protein [Homo sapiens]
HG1000271N 0_10000_gen e_prediction1	302	ldh_C ldh	0		(31-53)(80-102)	2	gi|27499559|ref|XP_062669.7|	similar to lactate dehydrogenase A
HG1000271N 0_160000_ge ne_prediction 1	302	ldh_C ldh	0			0	gi|27499559|ref|XP_062669.7|	similar to lactate dehydrogenase A
HG1000273N 0_160000_ge ne_prediction 1		no_pf am	0			0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000295N 0_160000_ge ne_prediction 1	465	no_pf am	0	0	(112-134)(149-171)(180-199)(204-226)	4	gi|12232393|ref|NP_073573.1|	hypothetical protein FLJ14153 [Homo
HG1000296N 0_160000_ge ne_prediction 1	912	no_pf am	0	0 (1-15)		0	gi|3327036|dbj|BAA31586.1|	KIAA0611 protein [Homo sapiens]
HG1000299N 0_160000_ge ne_prediction 1	459	no_pf am	0	0		0	gi|4503729|ref|NP_002005.1|	FK506-binding protein 4; FK506-binding
HG1000300N 0_10000_gen e_prediction1	1275	rvt pro_is omera se	0	0		0	gi|5052951|gb|AAD38785.1|AF 149422_2	unknown [Homo sapiens]
HG1000306N 0_0_gene_pre diction1		no_pf am	0	0 (14-29)		0		no_human_hit
HG1000306N 0_0_gene_pre diction2		no_pf am	0	0 (14-29)		0		no_human_hit

352

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000312N 0_160000_ge ne_prediction 1	173	no_pf am	0			0	gi|4506283|ref|NP_003454.1|	protein tyrosine phosphatase type IVA,
HG1000314N 0_1000_gene prediction1	167	no_pf am	0	(14-29)		0	gi|4506285|ref|NP_003470.1|	protein tyrosine phosphatase type
HG1000315N 0_160000_ge ne_prediction 1	415	Y_ph ospha tase	0			0	gi|4506291|ref|NP_002819.1|	protein tyrosine phosphatase,
HG1000330N 0_160000_ge ne_prediction 2	474	no_pf am	0			0	gi|10435451|dbj|BAB14590.1|	unnamed protein product [Homo sapiens]
HG1000330N 0_160000_ge ne_prediction	1962	no_pf am	0			0	gi|21702733|ref|NP_065898.1|	trinucleotide repeat containing 6;

353

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
4								
HG1000337N0_1000_gene_prediction1	263	Ribosomal_S4e KOW	0			0	gi|4506725|ref|NP_000998.1|	ribosomal protein S4, X-linked X
HG1000357N0_20000_gene_prediction1	3394	Zf_piccolo	0			0	gi|4522026|gb|AAD21789.1|	C-terminus matches KIAA0559, N-terminus
HG1000358N0_5000_gene_prediction1	1275	rvt	0			0	gi|2072948|gb|AAC51261.1|	putative p150 [Homo sapiens]
HG1000396N0_160000_gene_prediction2		no_pfam	0	(16-30)		0		no_human_hit
HG1000401N0_10000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1000414N0_160000_gene_prediction2		no_pfam	0			0		no_human_hit
HG1000416N0_160000_gene_prediction1	712	no_pfam	0	(1-17)		0	gi|1335205|emb|CAA36480.1|	ORFII [Homo sapiens]

354

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000428N0_160000_gene_prediction_1	1275	no_pfam	0			0	gi|2072951|gb|AAC51263.1|	putative p150 [Homo sapiens]
HG1000441N0_160000_gene_prediction_1		no_pfam	0			0		no_human_hit
HG1000441N0_160000_gene_prediction_2	1275	Exo_endo_phos	0	(1-23)		0	gi|5070622|gb|AAD39215.1|AF148856_2	unknown [Homo sapiens]
HG1000446N0_160000_gene_prediction_1	1275	rvt	0			0	gi|2072948|gb|AAC51261.1|	putative p150 [Homo sapiens]
HG1000446N0_160000_gene_prediction_2	1275	rvt	0			0	gi|2072953|gb|AAC51264.1|	putative p150 [Homo sapiens]
HG1000449N0_160000_gene_prediction_2	454	trypsin ldl_recept_a	0		(49-71)(166-188)	2	gi|13173471|ref|NP_076927.1|	transmembrane protease, serine 3
HG1000451N0_160000_gene_prediction	181	CSD	0			0	gi|17434360|ref|XP_065219.1|	similar to RNA-binding protein LIN-28

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
1								
HG1000455N0_10000_gen e_prediction1	1024	no_pf am	0			0	gi|20904033|ref|XP_139468.1|	similar to KIAA1662 protein [Homo
HG1000461N0_10000_gen e_prediction1		no_pf am	0			0		no_human_hit
HG1000476N0_1000_gene prediction1		no_pf am	0	(1-11)		0		no_human_hit
HG1000510N0_160000_ge ne prediction 1	42	no_pf am	0	(9-35)		0	gi|14249206|ref|NP_116044.1|	hypothetical protein MGC10997 [Homo
HG1000513N0_160000_ge ne prediction 1	248	GTP_EFTU_D2 GTP_EFTU GTP_EFTU_D3	0	0		0	gi|15928862|gb|AAH14892.1|A AH14892	Unknown (protein for
HG1000524N0_160000_ge ne prediction 1	465	no_pf am	0	(12-30)		0	gi|13959686|sp|P28476|GAR2_HUMAN	Gamma-aminobutyric-acid receptor

356

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000530N0_20000_gene_prediction1	358	no_pfam	0			0	gi|22042664|ref|XP_067445.5|	similar to KIAA1998 protein [Homo
HG1000530N0_160000_gene_prediction2		no_pfam	0			0		no_human_hit
HG1000534N0_160000_gene_prediction1	1275	rvt Transposase_22	0			0	gi|2072972|gb|AAC51276.1|	putative p150 [Homo sapiens]
HG1000545N0_160000_gene_prediction1	346	7tm_1	0		(52-74)(89-111)(132-154)(188-210)(223-241)(256-278)	6	gi|4885329|ref|NP_005295.1|	G protein-coupled receptor 41 [Homo
HG1000553N0_160000_gene_prediction1	473	no_pfam	0			0	gi|21750183|dbj|BAC03736.1|	unnamed protein product [Homo sapiens]
HG1000560N0_160000_gene_prediction2		no_pfam	0	(16-30)		0		no_human_hit
HG1000566N0_40000_gen_	42	no_pfam	0	(1-14)		0	gi|1424920|ref|NP_116044.1|	hypothetical protein MGC10997 [Homo

357

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
e_prediction1								
HG1000566N0_160000_gene_prediction1	42	no_pfam	0	(1-15)		0	gi|14249206|ref|NP_116044.1|	hypothetical protein MGC10997 [Homo
HG1000623N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1000624N0_160000_gene_prediction1	637	no_pfam	0	(1-11)		0	gi|17485449|ref|XP_066397.1|	similar to ebiP4655 [Anopheles gambiae
HG1000650N0_160000_gene_prediction1	121	Ribosomal_S17e DDE	0			0	gi|22049992|ref|XP_069827.2|	similar to 40S RIBOSOMAL PROTEIN S17
HG1000659N0_20000_gene_prediction1	550	no_pfam	0		(203-225)(235-257)(262-284)(288-310)(323-345)(355-372)(379-401)	7	gi|27498395|ref|XP_210662.1|	similar to hypothetical protein;
HG1000661N0_20000_gen	3394	Zf_piccolo	0			0	gi|4522026|gb|AAD21789.1|	C-terminus matches KIAA0559, N-terminus

358

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
e_prediction1								
HG1000664N0_160000_gene_prediction1	482	Hist_deacetyl	0			0	gi|13128860|ref|NP_004955.2|	histone deacetylase 1; reduced
HG1000690N0_20000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1000690N0_20000_gene_prediction2		no_pfam	0			0		no_human_hit
HG1000696N0_20000_gene_prediction1		no_pfam	0	(19-36)		0		no_human_hit
HG1000696N0_40000_gene_prediction1	367	no_pfam	0			0	gi|20809693|gb|AAH29202.1|	Similar to RIKEN cDNA 4933432E21 gene
HG1000697N0_160000_gene_prediction1	354	no_pfam	0			0	gi|4557435|ref|NP_001242.1|	CD68 antigen; Macrophage antigen CD68
HG1000704N0_160000_gene_prediction1	91	no_pfam	0			0	gi|27480400|ref|XP_211912.1|	hypothetical protein XP_211912 [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000711N0_20000_gene_prediction1	255	no_pfam	0	0 (1-22)		0	gi|7263962|emb|CAB81646.1|bA395L14.12	(novel protein similar to
HG1000740N0_10000_gene_prediction1	719	no_pfam	0	0		0	gi|20864261|ref|XP_143091.1|	similar to hypothetical protein
HG1000743N0_160000_gene_prediction1	411	no_pfam	0	0		0	gi|27480224|ref|XP_166599.2|	similar to hypothetical protein
HG1000781N0_160000_gene_prediction1	325	no_pfam	0	0		0	gi|14290590|gb|AAH09074.1|AAH09074	Similar to CGI-62 protein [Homo
HG1000781N0_160000_gene_prediction2	1275	no_pfam	0	0		0	gi|2072948|gb|AAC51261.1|	putative p150 [Homo sapiens]
HG1000788N0_1000_gene_prediction1	820	LRRCT	0	0 (1-23)	(418-440)	1	gi|16168273|ref|XP_056282.1|	similar to KIAA1904 protein [Homo
HG1000808N0_160000_gene_prediction1		no_pfam	0	0		0		no_human_hit
HG1000817N0_160000_gene_prediction	646	HSP70	0	0		0	gi|5729877|ref|NP_006588.1|	heat shock 70kDa protein 8 isoform 1;

360

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000822N0_20000_gene_prediction1	482	Hist_deacetyl	0			0	gi|13128860|ref|NP_004955.2|	histone deacetylase 1; reduced
HG1000842N0_160000_gene_prediction1	1275	rvt	0			0	gi|2072977|gb|AAC51279.1|	putative p150 [Homo sapiens]
HG1000842N0_160000_gene_prediction2		rvt rnaseH Gag_p30	0			0		no_human_hit
HG1000870N0_160000_gene_prediction1	591	no_pfam	0		(53-75)(90-112)(125-144)	3	gi|27686465|ref|XP_237439.1|	similar to organic anion transporter
HG1000870N0_160000_gene_prediction2	591	no_pfam	0		(53-75)(90-112)(125-144)	3	gi|27686465|ref|XP_237439.1|	similar to organic anion transporter
HG1000878N0_20000_gene_prediction1	591	OATP_N	0		(53-75)(85-107)(137-159)	3	gi|27686465|ref|XP_237439.1|	similar to organic anion transporter

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000878N0_20000_gene_prediction2	591	OATP_N	0		(53-75)(85-107)(137-159)	3	gi|27686465|ref|XP_237439.1|	similar to organic anion transporter
HG1000906N0_5000_gene_prediction1	1211	Armadillo_seg	0			0	gi|4505843|ref|NP_003619.1|	plakophilin 4 [Homo sapiens]
HG1000906N0_160000_gene_prediction2	1275	Armadillo_seg rvt Transposase_22 Exo_endo_phos	0			0	gi|5052951|gb|AAD38785.1|AF149422_2	unknown [Homo sapiens]
HG1000910N0_160000_gene_prediction1	395	no_pfam	0		(112-130)(153-175)	2	gi|24431977|ref|NP_060523.2|	hypothetical protein FLJ10307 [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000948N0_160000_gene_prediction1	1275	Exo_endo_phos	0			0	gi|5070622|gb|AAD39215.1|AF148856_2	unknown [Homo sapiens]
HG1000955N0_160000_gene_prediction1	360	inositol_P_DUF803	0			0	gi|24308307|ref|NP_112184.1|	hypothetical protein MGC5466 [Homo
HG1000959N0_160000_gene_prediction1	253	transmembrane4	0		(98-120)(212-234)	2	gi|21237748|ref|NP_004348.2|	CD151 antigen, isoform a;
HG1000959N0_5000_gene_prediction1	253	transmembrane4	0		(53-75)(85-107)(199-221)	3	gi|21237748|ref|NP_004348.2|	CD151 antigen, isoform a;
HG1000990N0_5000_gene_prediction1	234	no_pfam	0	(1-18)		0	gi|8924262|ref|NP_061113.1|	triggering receptor expressed on
HG1000994N0_10000_gene_prediction1	441	no_pfam	0	(1-19)		0	gi|21361800|ref|NP_478067.2|	chromosome 21 open reading frame 63
HG1000994N0_160000_gene_prediction2	869	7tm_1	0		(111-133)	1	gi|27498059|ref|XP_068166.3|	similar to olfactory receptor MOR145-2

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000994N 0_10000_gen e_prediction2	441	no_pf am	0	(1-19)		0	gi|2136180|ref|NP_478067.2|	chromosome 21 open reading frame 63
HG1001001N 0_160000_ge ne_prediction 1	477	7tm_ 1	0		(121-143)(155-177)(192-214)(260-282)	4	gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor
HG1001001N 0_0_gene_pre diction1	869	DUF2 70	0		(49-71)(92-111)(121-143)(164-186)(196-215)(312-334)(382-404)	7	gi|27498059|ref|XP_068166.3|	similar to olfactory receptor MOR145-2
HG1001002N 0_160000_ge ne_prediction 1	277	ig	0			0	gi|36945|emb|CAA26435.1|	TCR-alpha chain [Homo sapiens]

364

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001003N 0_0_gene_pre diction1	869	DUF2 70	0		(49-71)(92-111)(121-143)(164-186)(196-215)(312-334)(382-404)	7	gi|27498059|ref|XP_068166.3|	similar to olfactory receptor MOR145-2
HG1001007N 0_160000_ge ne_prediction 2	192	no_pf am	0			0	gi|13375791|ref|NP_078871.1|	hypothetical protein FLJ12666 [Homo
HG1001011N 0_160000_ge ne_prediction 1	472	TMS _TDE	0		(84-106)(119-138)(148-170)(246-268)	4	gi|6382026|dbj|BAA86567.1|	KIAA1253 protein [Homo sapiens]
HG1001011N 0_160000_ge ne_prediction 2	508	Band _41	0			0	gi|25047957|ref|XP_130582.2|	similar to hypothetical protein

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001014N 0_160000_ge ne_prediction 1	477	7tm_ 1	0		(121-143)(155-177)(192-214)(260-282)	4	gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor
HG1001014N 0_5000_gene _prediction1	477	7tm_ 1	0		(121-143)(155-177)(192-214)(260-282)	4	gi|4503391|ref|NP_000789.1|	dopamine receptor D5; dopamine receptor
HG1001017N 0_160000_ge ne_prediction 1		no_pf am	0			0		no_human_hit
HG1001020N 0_160000_ge ne_prediction 1		no_pf am	0			0		no_human_hit
HG1001024N 0_160000_ge ne_prediction 1	667	no_pf am	0	(1-19)	(366-388)(395-417)(432-454)(466-488)(503-520)	5	gi|21749264|dbj|BAC03564.1|	unnamed protein product [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
					520)			
HG1001024N0_160000_gene_prediction2	667	no_pfam		0 (1-19)	(366-388)(395-417)(432-454)(466-488)(503-520)	5	gi|21749264|dbj|BAC03564.1|	unnamed protein product [Homo sapiens]
HG1001031N0_160000_gene_prediction1	360	no_pfam		0	(117-139)	1	gi|17444989|ref|XP_059461.1|	similar to hypothetical protein
HG1001035N0_5000_gene prediction1	262	no_pfam		0		0	gi|21755758|dbj|BAC04753.1|	unnamed protein product [Homo
HG1001043N0_160000_gene_prediction1	555	MRVI1		0		0	gi|5453724|ref|NP_006143.1|	lymphoid-restricted membrane protein
HG1001046N0_5000_gene prediction1	107	no_pfam		0		0	gi|18553203|ref|XP_098065.1|	hypothetical protein XP_098065 [Homo
HG1001046N0_160000_ge	113	no_pfam		0		0	gi|27478106|ref|XP_211857.1|	hypothetical protein XP_211857 [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
ne_prediction 1								
HG1001047N 0_1000_gene prediction1	408	helica se_C	0			0	gi|673433|emb|CAA40268.1|	protein synthesis initiation factor 4A
HG1001048N 0_160000_ge ne_prediction 1	566	SSF	0		(104-126)(183-205)(220-242)(249-271)(291-313)(350-372)	6	gi|24980856|gb|AAH39868.1|	Unknown (protein for MGC:49034) [Homo
HG1001048N 0_160000_ge ne_prediction 2	200	Ran_BP1	0	(7-25)		0	gi|938026|dbj|BAA07269.1|	Ran-binding protein 1 [Homo sapiens]
HG1001144N 0_20000_gen e_prediction1	338	Trans posas e 22	0			0	gi|5070621|gb|AAD39214.1|AF 148856_1	unknown [Homo sapiens]
HG1001148N 0_160000_ge ne_prediction 2	814	disint egrin	0		(290-312)	1	gi|15778976|gb|AAH14566.1|A AH14566	a disintegrin and
HG1001172N 0_160000_ge ne_prediction	150	WD4 0	0			0	gi|18544733|ref|XP_087052.1|	similar to hypothetical protein

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
1								
HG1001172N0_20000_gene_prediction1	815	no_pfam	0			0	gi|21542539|gb|AAH33043.1|	echinoderm microtubule associated
HG1001194N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1001223N0_160000_gene_prediction1	1275	rvt	0			0	gi|5070622|gb|AAD39215.1|AF148856_2	unknown [Homo sapiens]
HG1001284N0_160000_gene_prediction1	685	no_pfam	0			0	gi|27485931|ref|XP_084736.6|	similar to CG5815-PA [Drosophila]
HG1001284N0_160000_gene_prediction2	685	no_pfam	0			0	gi|27485931|ref|XP_084736.6|	similar to CG5815-PA [Drosophila]
HG1001292N0_160000_gene_prediction1	226	vATP-synt_E	0			0	gi|4502317|ref|NP_001687.1|	ATPase, H+ transporting, lysosomal
HG1001302N0_160000_gene_prediction1	1734	no_pfam	0			0	gi|14149671|ref|NP_055927.1|	microtubule associated testis specific

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
1								
HG1001323N 0_160000_ge ne_prediction 1		no_pf am	0			0		no_human_hit
HG1001328N 0_5000_gene _prediction1		no_pf am	0			0		no_human_hit
HG1001328N 0_40000_gen e_prediction1	1275	rvt	0			0	gi|2072977|gb|AAC51279.1|	putative p150 [Homo sapiens]
HG1001331N 0_0_gene_pre diction1	344	no_pf am	0		(179-201)(222-241)(251-273)	3	gi|18573404|ref|XP_088565.1|	similar to brain protein [Homo
HG1001348N 0_160000_ge ne_prediction 1		no_pf am	0			0		no_human_hit
HG1001349N 0_160000_ge ne_prediction 1	135	no_pf am	0			0	gi|27671653|ref|XP_213311.1|	similar to KIAA1639 protein [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001354N0_160000_gene_prediction1	1427	homeobox	0			0	gi|20542050|ref|XP_033370.6|	similar to KIAA1762 protein [Homo
HG1001361N0_160000_gene_prediction1	797	no_pfam	0		(55-77)(92-114)	2	gi|4759136|ref|NP_004202.1|	solute carrier family 6
HG1001376N0_160000_gene_prediction1	1353	no_pfam	0			0	gi|7512821|pir||T00347	hypothetical protein DKFZp566G1246.1,
HG1001376N0_5000_gene_prediction1	1353	no_pfam	0			0	gi|7512821|pir||T00347	hypothetical protein DKFZp566G1246.1,
HG1001376N0_20000_gene_prediction1	1353	no_pfam	0			0	gi|7512821|pir||T00347	hypothetical protein DKFZp566G1246.1,
HG1001376N0_5000_gene_prediction2	1353	no_pfam	0			0	gi|7512821|pir||T00347	hypothetical protein DKFZp566G1246.1,
HG1001376N0_5000_gene_prediction3	1320	no_pfam	0			0	gi|22052701|ref|XP_051956.8|	similar to hypothetical protein
HG1001436N0_5000_gene_prediction1		no_pfam	0			0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001436N 0_20000_gen e_prediction1	720	no_pf am	0			0	gi|13376064|ref|NP_079017.1|	hypothetical protein FLJ12827 [Homo
HG1001436N 0_160000_ge ne_prediction 1		no_pf am	0			0		no_human_hit
HG1001484N 0_160000_ge ne_prediction 1	1275	rvt Transposas e_22	0			0	gi|20729064|gb|AAC51271.1|	putative p150 [Homo sapiens]
HG1001500N 0_160000_ge ne_prediction 1	294	Nucle oplas min	0			0	gi|18314408|gb|AAH21983.1|A AH21983	nucleophosmin (nucleolar
HG1001500N 0_160000_ge ne_prediction 2	224	Trans posas e_22	0			0	gi|27484907|ref|XP_210358.1|	similar to p40 [Homo sapiens]
HG1001508N 0_160000_ge ne_prediction 1	892	no_pf am	0	(11-27)		0	gi|27705266|ref|XP_228160.1|	similar to hypothetical protein
HG1000898N 0_5000_gene _prediction1	597		0			0	gi|27462211|gb|AAO15382.1|	multiple hat domains [Homo sapiens]

Examples

[0604] The examples, which are intended to be purely exemplary of the invention and should therefore not be considered to limit the invention in any way, also describe and detail aspects and embodiments of the invention discussed above. The examples are not intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

[0605] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications can be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

[0606] Additional objects and advantages of the invention will be set forth in part in the description which follows, and in part will be obvious from the description, or may be learned by practice of the invention. The objects and advantages of the invention will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. Moreover, advantages described in the body of the specification, if not included in the claims, are not per se limitations to the claimed invention.

[0607] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed. Moreover, it must be understood that the invention is not limited to the particular embodiments described, as such may, of course, vary. Further, the terminology used to describe particular embodiments is not intended to be limiting, since the scope of the present invention will be limited only by its claims.

[0608] With respect to ranges of values, the invention encompasses each intervening value between the upper and lower limits of the range to at least a tenth of the lower limit's unit, unless the context clearly indicates otherwise. Further, the

invention encompasses any other stated intervening values. Moreover, the invention also encompasses ranges excluding either or both of the upper and lower limits of the range, unless specifically excluded from the stated range.

[0609] Unless defined otherwise, the meanings of all technical and scientific terms used herein are those commonly understood by one of ordinary skill in the art to which this invention belongs. One of ordinary skill in the art will also appreciate that any methods and materials similar or equivalent to those described herein can also be used to practice or test the invention. Further, all publications mentioned herein are incorporated by reference.

[0610] It must be noted that, as used herein and in the appended claims, the singular forms "a," "or," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a subject polypeptide" includes a plurality of such polypeptides and reference to "the agent" includes reference to one or more agents and equivalents thereof known to those skilled in the art, and so forth.

[0611] Further, all numbers expressing quantities of ingredients, reaction conditions, % purity, polypeptide and polynucleotide lengths, and so forth, used in the specification and claims, are modified by the term "about," unless otherwise indicated. Accordingly, the numerical parameters set forth in the specification and claims are approximations that may vary depending upon the desired properties of the present invention. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits, applying ordinary rounding techniques. Nonetheless, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors from the standard deviation of its experimental measurement.

[0612] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

Example 1 Expression in *E. coli*

[0613]   Sequences can be expressed in *E. coli*. Any one or more of the sequences according to SEQ ID NOS.: 1 -1231 and 2463 - 3697 can be expressed in *E. coli* by subcloning the entire coding region, or a selected portion thereof, into a prokaryotic expression vector. For example, the expression vector pQE16 from the QIA expression prokaryotic protein expression system (Qiagen, Valencia, CA) can be used. The features of this vector that make it useful for protein expression include an efficient promoter (phage T5) to drive transcription, expression control provided by the lac operator system, which can be induced by addition of IPTG (isopropyl-beta-D-thiogalactopyranoside), and an encoded 6XHis tag coding sequence. The latter is a stretch of six histidine amino acid residues which can bind very tightly to a nickel atom. This vector can be used to express a recombinant protein with a 6XHis. tag fused to its carboxyl terminus, allowing rapid and efficient purification using Ni-coupled affinity columns.

[0614]   The entire or the selected partial coding region can be amplified by PCR, then ligated into digested pQE16 vector. The ligation product can be transformed by electroporation into electrocompetent *E. coli* cells (for example, strain M15[pREP4] from Qiagen), and the transformed cells may be plated on ampicillin-containing plates. Colonies may then be screened for the correct insert in the proper orientation using a PCR reaction employing a gene-specific primer and a vector-specific primer. Also, positive clones can be sequenced to ensure correct orientation and sequence. To express the proteins, a colony containing a correct recombinant clone can be inoculated into L-Broth containing 100 µg/ml of ampicillin, and 25 µg/ml of kanamycin, and the culture allowed to grow overnight at 37 degrees C. The saturated culture may then be diluted 20-fold in the same medium and allowed to grow to an optical density of 0.5 at 600 nm. At this point, IPTG can be added to a final concentration of 1 mM to induce protein expression.. After growing the culture for an additional 5 hours, the cells may be harvested by centrifugation at 3000 times *g* for 15 minutes.

[0615]   The resultant pellet can be lysed with a mild, nonionic detergent in 20 mM Tris HCl (pH 7.5) (B PER.TM. Reagent from Pierce, Rockford, IL), or by sonication until the turbid cell suspension turns translucent. The resulting lysate can be further purified using a nickel-containing column (Ni-NTA spin column from

Qiagen) under non-denaturing conditions. Briefly, the lysate will be adjusted to 300 mM NaCl and 10 mM imidazole, then centrifuged at 700 times g through the nickel spin column to allow the His-tagged recombinant protein to bind to the column. The column will be washed twice with wash buffer (for example, 50 mM $NaH_2 PO_4$, pH 8.0; 300 mM NaCl; 20 mM imidazole) and eluted with elution buffer (for example, 50 mM NaH2 PO4, pH 8.0; 300 mM NaCl; 250 mM imidazole). All the above procedures will be performed at 4 degrees C. The presence of a purified protein of the predicted size can be confirmed with SDS-PAGE.

Example 2: Expression in Mammalian Cells

[0616] The sequences encoding the proteins of Example 1 can be cloned into the pENTR vector (Invitrogen) by PCR and transferred to the mammalian expression vector pDEST12.2 per manufacturer's instructions (Invitrogen). Introduction of the recombinant construct into the host cell can be effected by transfection with Fugene 6 (Roche) per manufacturer's instructions. The host cells containing one of polynucleotides of the invention can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF). A number of types of cells can act as suitable host cells for expression of the proteins. Mammalian host cells include, for example, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells.

Example 3: Expression in Cell-Free Translation Systems

[0617] Cell-free translation systems can also be employed to produce proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors containing SP6 or T7 promoters for use with prokaryotic and eukaryotic hosts have been described (Sambrook et al., 1989). These DNA constructs can be used to produce proteins in a rabbit reticulocyte lysate system or in a wheat germ extract system.

[0618] Specific expression systems of interest include plant, bacterial, yeast, insect cell and mammalian cell derived expression systems. Expression systems in plants include those described in U.S. Patent No. 6,096,546 and U.S. Patent No. 6,127,145. Expression systems in bacteria include those described by Chang et al.,

1978, Goeddel et al., 1979, Goeddel et al., 1980, EP 0 036,776, U.S. Patent No. 4,551,433; DeBoer et al., 1983, and Siebenlist et al., 1980.

[0619]   Mammalian expression is further accomplished as described in Dijkema et al. 1985, Gorman et al., 1982, Boshart et al., 1985, and U.S. Patent No. 4,399,216. Other features of mammalian expression are facilitated as described in Ham and Wallace, Meth. Enz., 1979, Barnes and Sato, 1980, U.S. Patent Nos. 4,767,704, 4,657,866, 4,927,762, 4,560,655, WO 90/103430, WO 87/00195, and U.S. RE 30,985.

Example 4: Expression of the Secreted Factors in Yeast

[0620]   Primers can be designed to amplify the secreted factors using PCR and cloned into pENTR/D-TOPO vectors (Invitrogen, Carlsbad, CA). The secreted factors in pENTR/D-TOPO can be cloned into the yeast expression vector pYES-DEST52 by Gateway LR reaction (Invitrogen, Carlsbad, CA). The resulting yeast expression vectors can be transformed into INVSc1 strain from Invitrogen to express the secreted factors according to the manufacturer's protocol (Invitrogen, Carlsbad CA). The expressed secreted factors will have a 6XHis tag at the C-terminal. Expressed protein can be purified with ProBond™ resin (Invitrogen, Carlsbad, CA).

[0621]   Expression systems in yeast include those described in Hinnen et al., 1978, Ito et al., 1983, Kurtz et al., 1986, Kunze et al., 1985, Gleeson et al., 1986, Roggenkamp et al., 1986, Das et al., 1984, De Louvencourt et al., 1983, Van den Berg et al., 1990, Kunze et al., 1985, Cregg et al. 1985, U.S. Patent No. 4,837,148, U.S. Patent No. 4,929,555, Beach and Nurse, 1981, Davidow et al., 1985, Gaillardin et al., 1985, Ballance et al., 1983, Tilburn et al., 1983, Yelton et al., 1984, Kelly and Hynes, 1985, EP 0 244,234, and WO 91/00357.

Example 5: Expression of Secreted Factors in Baculovirus Expression System.

[0622]   The secreted factors in pENTR/D-TOPO can be cloned into Baculovirus expression vector pDEST10 by Gateway LR reaction (Invitrogen, Carlsbad, CA). The secreted factors can be expressed by the Bac-to-Bac expression system from Invitrogen (Carlsbad CA), briefly described as follows. The expression vectors containing the secreted factors are transformed into competent DH10Bac™ E. coli strain and selected for transposition. The resulting E coli contain recombinant bacmid that contains the secreted factor. High molecular weight DNA can be isolated from the E. coli containing the recombinant bacmid and then transfected into insect

cells with Cellfectin reagent. The expressed secreted factors will have a 6XHis tag at N-terminal. Expressed protein will be purified by ProBond™ resin (Invitrogen, Carlsbad, CA).

[0623]   Expression of heterologous genes in insects can be accomplished as described in U.S. Patent No. 4,745,051; Doerfler et al., 1087; Friesen *et al.*, 1986; EP 0 127,839, EP 0 155,476, Vlak *et al.*, 1988, Miller *et al.*, 1988, Carbonell *et al.*, 1988, Maeda *et al.*, 1985, Lebacq-Verheyden *et al.*, 1988, Smith *et al.*, 1985, Miyajima *et al.*; and Martin *et al.*, 1988. Numerous baculoviral strains and variants and corresponding permissive insect host cells from hosts have been previously described (Setlow et al., 1986, Luckow *et al.*, 1988; Miller *et al.*, 1986; Maeda *et al.*, 1985).

Example 6: Primer Design

[0624]   To design the forward primer for PCR amplification, the melting point of the first 20 to 24 bases of the primer can be calculated by counting total A and T residues, then multiplying by 2. To design the reverse primer for PCR amplification, the melting point of the first 20 to 24 bases of the reverse complement, with the sequences written from 5-prime to 3-prime can be calculated by counting the total G and C residues, then multiplying by 4. Both start and stop codons can be present in the final amplified clone. The length of the primers is such to obtain melting temperatures within 63 degrees C to 68 degrees C. Adding the bases "CACC" to the forward primer renders it compatible for cloning the PCR product with the TOPO pENTR/D (Invitrogen, CA).

Example 7: Reverse Transcriptase Reaction

[0625]   cDNA can be prepared by the following method. Between 200 ng and 1.0 µg mRNA is added to 2 µl DMSO and the volume adjusted to 11 µl with DEPC-treated water. One µl Oligo dT is added to the tube, and the mixture is heated at 70° C for 5 min., quickly chilled on ice for 2 min., and the mixture is collected at the bottom of the tube by brief centrifugation. The following 1st strand components are then added to the mRNA mixture: 2 µl 10X Stratascript (Stratagene, CA) 1st strand buffer, 1 µl 0.1 M DTT, 1 µl 10 mM dNTP mix (10 mM each of dG, dA, dT and dCTP), 1 µl RNAse inhibitor, 3 µl Stratascript RT (50 U/ µl). The contents are gently mixed and the mixture collected by brief centrifugation. The mixture is incubated in a 42° C water bath for 1 hour, placed in a 70° C water bath for 15 min. to stop the reaction, transferred to ice for 2 min., and centrifuged briefly in a microfuge to collect the reaction product at the bottom of the reaction vessel. Two µl RNAse H is then

added to the tube, the contents are mixed well, incubated at 37° C in a water bath for 20 min., and centrifuged briefly in a microfuge to collect the reaction product at the bottom of the reaction vessel. The reaction mixture can proceed directly to PCR or be stored at – 20° C.

Example 8: Full Length PCR

[0626]   Full length PCR can be achieved by placing the products of the reaction described in Example 7, with primers diluted to 5μM in water, into a reaction vessel and adding a reaction mixture composed of 1x Taq buffer, 25 mM dNTP, 10 ng cDNA pool, TaqPlus (Stratagene, CA) (5u/ul), PfuTurbo (Stratagene, CA) (2.5u/ul), water. The contents of the reaction vessel are then mixed gently by inversion 5-6 times, placed into a reservoir where 2μl F_1/R_1 primers are added, the plate sealed and placed in the thermocycler. The PCR reaction is comprised of the following eight steps. Step 1: 95° C for 3 min. Step 2: 94° C for 45 sec. Step 3: 0.5° C/sec to 56-60° C. Step 4: 56-60° C for 50 sec. Step 5: 72° C for 5 min. Step 6: Go to step 2, perform 35-40 cycles. Step 7: 72° C for 20 min. Step 8: 4° C.

[0627]   The products can then be separated on a standard 0.8 to 1.0% agarose gel at 40 to 80 V, the bands of interest excised by cutting from the gel, and stored at – 20° C until extraction. The material in the bands of interest can be purified with QIAquick 96 PCR Purification Kit (Qiagen, CA) according to the manufacturer instructions. Cloning can be performed with the Topo Vector pENTR/D-TOPO vector (Invitrogen, CA) according to the manufacturer's instructions.

References

[0628]   The specification is most thoroughly understood in light of the following references, all of which are hereby incorporated by reference in their entireties. The disclosures of the patents and other references cited above are also hereby incorporated by reference.

1. Agou, F., Quevillon, S., Kerjan, P., Latreille, M.T., Mirande, M. (1996) Functional replacement of hamster lysyl-tRNA synthetase by the yeast enzyme requires cognate amino acid sequences for proper tRNA recognition. *Biochemistry* 35:15322-15331.

2. Agrawal, S., Crooke, S.T. eds. (1998) Antisense Research and Application (Handbook of Experimental Pharmacology, Vol 131). Springer-Verlag New York, Inc.

3. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D. (1994) Molecular Biology of the Cell. 3rd ed. Garland Publishing, Inc.

4. Alexander, D.R. (2000) The CD45 tyrosine phosphatase: a positive and negative regulator of immune cell function. *Semin. Immunol* 12:349-359.

5. Allison, A.C. (2000) Immunosuppressive drugs: the first 50 years and a glance forward. *Immunopharmacology* 47:63-83.

6. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic alignment search tool. *J. Mol. Biol.* 215:403-410.

7. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zheng, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.

8. Amor, J.C., Harrison, D.H., Kahn, R.A., Ringe, D. (1994) Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature* 372:704-708.

9. Andreeff, M., Pinkel, D. eds. (1999) Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications. John Wiley & Sons.

10. Andres, D.A., Shao, H., Crick, D.C., Finlin, B.S. (1997) Expression cloning of a novel farnesylated protein, RDJ2, encoding a DnaJ protein homologue. *Arch. Biochem. Biophys.* 346:113-124.

11. Ansel, H.C., Allen, L., Popovich, N.G. eds. (1999) Pharmaceutical Dosage Forms and Drug Delivery Systems. 7th ed. Lippencott Williams and Wilkins Publishers.

12. Aubry, M., Marineau, C., Zhang, F.R., Zahed, L., Figlewicz, D., Delattre, O., Thomas, G., de Jong, P.J., Julien, J.P., Rouleau, G.A. (1992) Cloning of six new genes with zinc finger motifs mapping to short and long arms of human acrocentric chromosome 22 (p and q11.2). *Genomics* 13:641-648.

13. Ausubel, F., Brent. R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., eds. (1999) Short Protocols in Molecular Biology. 4th ed. Wiley & Sons.

14. Baksh, S., Burakoff, S.J. (2000) The role of calcineurin in lymphocyte activation. *Semin. Immunol.* 12:405-415.

15. Ballance, D.J., Buxton, F.P., Turner, G. (1983) Transformation of *Aspergillus nidulans* by the orotidine-5'-phosphate decarboxylase gene of *Neurospora crassa. Biochem. Biophys. Res. Commun.* 112:284-289.

16. Barany, F. (1985) Single-stranded hexameric linkers: a system for in-phase insertion mutagenesis and protein engineering. *Gene* 37:111-123.

17. Barnes, D., Sato, G. (1980) Methods for growth of cultured cells in serum-free medium. *Anal. Biochem.* 102:255-270.

18. Barton, M.C., Hoekstra, M.F., Emerson, B.M. (1990) Site-directed, recombination-mediated mutagenesis of a complex gene locus. *Nucleic Acids Res.* 18:7349-7355.

19. Bashkin, J.K., Sampath, U., Frolova, E. (1995) Ribozyme mimics as catalytic antisense reagents. *Appl. Biochem. Biotechnol.* 54:43-56.

20. Bassett, D.E., Eisen, M.B., Boguski, M.S. (1999) Gene expression informatics - it's all in your mine. *Nature Genetics* 21:51-55.

21. Bast, R.C., Kufe, D.W., Pollock, R.E., Weichselbaum, R.R., Holland, J.F., Frei, E., eds. (2000) Cancer Medicine. 5th ed. B.C. Decker, Inc.

22. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., Sonnhammer, E.L.L. (2000) *Nucleic Acids Research* 30:276-280.

23. Battini, R., Ferrari, S., Kaczmarek, L., Calabretta, B., Chen, S.T., Baserga, R. (1987) Molecular cloning of a cDNA for a human ADP/ATP carrier which is growth-regulated. *J. Biol. Chem.* 262:4355-4359.

24. Beach, D., Durkacz, B., Nurse, P. (1982) Functionally homologous cell cycle control genes in budding and fission yeast. *Nature* 300:706-709.

25. Beigelman, L., Karpeisky, A., Matulic-Adamic, J., Haeberli, P., Sweedler, D., Usman, N. (1995) Synthesis of 2'-modified nucleotides and their incorporation into hammerhead ribozymes. *Nucleic Acids Res.* 23:4434-4442.

26. Bennett, J. (2000) Gene therapy for retinitis pigmentosa. *Curr. Opin. Mol. Ther.* 2:420-425.

27. Berinstein, N.L. (2002) Carcinoembryonic antigen as a target for therapeutic anticancer vaccines: a review. *J. Clin. Oncol.* 20:2197-2207.

28. Bibikova, M., Beumer, K., Trautman, J.K., Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* 300:764.

29. Birney, E., Durbin, R. (2000) Using GeneWise in the *Drosophila* annotation experiment. *Genome Res.* 10:547-548.

30. Blackwell, J.M., Barton, C.H., White, J.K., Searle, S., Baker, A.M., Williams, H., Shaw, M.A. (1995) Genomic organization and sequence of the human NRAMP gene: identification and mapping of a promoter region polymorphism. *Mol. Med.* 1:194-205.

31. Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Porsch-Ozcurumez, M., Kaminski, W.E., Hahmann, H.W., Oette, K., Rothe, G., Aslanidis, C., Lackner, K.J., Schmitz, G. (1999) The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat. Genet.* 1999 22:347-351.

32. Bonifaci, N., Moroianu, J., Radu, A., Blobel, G. (1997) Karyopherin beta2 mediates nuclear import of a mRNA binding protein. *Proc. Natl. Acad. Sci.* 94:5055-5060.

33. Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B., Schaffner, W. (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* 41:521-530.

34. Bowtell, D.D.L. (1999) Options available - from start to finish - for obtaining expression data by microarray. *Nature Genetics* 21:25-32.

35. Brenner, S., Williams, S.R., Vermass, E.H., Storck, T., Moon, K., McCollum, C., Mao, J.I., Luo, S., Kirchner, J.J., Eletr, S., DuBridge, R.B., Burcham, T., Albrecht, G. (2000) *In vitro* cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci. USA* 97:1665-1670.

36. Brock, G. (2000) Sildenafil citrate (Viagra®). *Drugs Today* 36:125-134.

37. Brown, J.R., Daar, I.O., Krug, J.R., Maquat, L.E. (1985) Characterization of the functional gene and several processed pseudogenes in the human triosephosphate isomerase gene family. *Mol. Cell Biol.* 5:1694-1706.

38. Brown, P.O, Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nature Genetics* 21:33-37.

39. Brunelleschi, S., Penengo, L., Santoro, M.M., Gaudino, G. (2002) Receptor tyrosine kinases as target for anti-cancer therapy. *Curr. Pharm. Des.* 8:1959-1972.

40. Brutlag, D.L., Dautricourt, J.P., Diaz, R., Fier, J., Moxon, B., Stamm, R. (1993). BLAZE: An implementation of the Smith-Waterman comparison algorithm on a massively parallel computer. *Computers and Chemistry* 17:203-207.

41. Carbonell, L.F., Hodge, M.R., Tomalski, M.D., Miller, L.K. (1988) Synthesis of a gene coding for an insect-specific scorpion neurotoxin and attempts to express it using baculovirus vectors. *Gene* 73:409-418.

42. Chakravarty, A. (1999) Population genetics - making sense out of sequence. *Nature Genetics* 21:56-60.

43. Chalifour, L.E., Fahmy, R., Holder, E.L., Hutchinson, E.W., Osterland, C.K., Schipper, H.M., Wang, E. (1994) A method for analysis of gene expression patterns. *Anal. Biochem.* 216: 299-304.

44. Chalut, C., Gallois, Y., Poterszman, A., Moncollin, V., Egly, J.M. (1995) Genomic structure of the human TATA-box-binding protein (TBP). *Gene* 161:277-282.

45. Chang, A.C., Nunberg, J.H., Kaufman, R.J., Erlich, H.A., Schimke, R.T., Cohen, S.N. (1978) Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. *Nature* 275:617-624.

46. Chang, M.S., Chang, C.L., Huang, C.J., Yang, Y.C. (2000) p29, a novel GCIP-interacting protein, localizes in the nucleus. *Biochem. Biophys. Res. Commun.* 279:732-737.

47. Chen, F.W., Ioannou, Y.A. (1998) Ribosomal proteins in cell proliferation and apoptosis. *Int. Rev. Immunol.* 18:429-448.

48. Chen, S.Y., Bagley, J., Marasco, W.A. (1994) Intracellular antibodies as a new class of therapeutic molecule for gene therapy. *Hum. Gene Ther.* 5:595-601.

383

49. Cheng, W.F., Hung, C.F., Chai, C.Y., Hsu, K.F., He, L., Ling, M., Wu, T.C. (2001) Tumor-specific immunity and angiogenesis generated by a DNA vaccine encoding calreticulin linked to a tumor antigen. *J. Clin. Invest.* 108:669-678.

50. Cheung, V.G., Morley, M., Aquilar, F., Massimi, A., Kucherlapati, R., Childs, G. (1999) Making and reading microarrays. *Nature Genetics* 21:15-19.

51. Chien, C., Bartel, P.L., Sternglanz, R., Fields S. (1991) The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci.* 88:9578-9581.

52. Christa, L., Simon, M.T., Flinois, J.P., Gebhardt, R., Brechot, C., Lasserre, C. (1994) Overexpression of glutamine synthetase in human primary liver cancer. *Gastroenterology* 106:1312-1320.

53. Clark, C.M., Karlawish, J.H. (2003) Alzheimer disease: current concepts and emerging diagnostic and therapeutic strategies. *Ann. Intern. Med.* 138:400-410.

54. Coffin, J.M., Hughes, S.H., Varmus, H.E. (1997) Retroviruses. Cold Spring Harbor Laboratory Press.

55. Cole, K.A., Krizman, D.B., Emmert-Buck, M.R. (1999) The genetics of cancer - a 3D model. *Nature Genetics* 21:38-41.

56. Colicelli, J., Lobel, L.I., Goff, S.P. (1985) A temperature-sensitive mutation constructed by "linker insertion" mutagenesis. *Mol. Gen. Genet.* 199:537-539.

57. Collins, F.S. (1999) Microarrays and macroconsequences. *Nature Genetics* 21:2.

58. Comuzzie, A.G., Allison, D.B. (1998) The search for human obesity genes. *Science* 280:1374-1377.

59. Cormand, B., Montfort, M., Chabas, A., Vilageliu, L., Grinberg, D. (1997) Genetic fine localization of the beta-glucocerebrosidase (GBA) and prosaposin (PSAP) genes: implications for Gaucher disease. *Hum. Genet.* 100:75-79.

60. Cregg, J.M., Barringer, K.J., Hessler, A.Y., Madden, K.R. (1985) *Pichia pastoris* as a host system for transformations. *Mol. Cell. Biol.* 5:3376-3385.

61. Crooke, S.T. (1996) Progress in antisense therapeutics. *Med. Res. Rev.* 16:319-344.

62. Crouch, R.J. (1990) Ribonuclease H: from discovery to 3D structure. *New Biol.* 2:771-777.

63. Curcio, L.D., Bouffard, D.Y., Scanlon, K.J. (1997) Oligonucleotides as modulators of cancer gene expression. *Pharmacol. Ther.* 74:317-332.

64. Das, S., Kellermann, E., Hollenberg, C.P. (1984) Transformation of *Kluyveromyces fragilis*. *J. Bacteriol.* 158:1165-1167.

65. Davidow, L.S., Kaczmarek, F.S., DeZeeuw, J.R., Conlon, S.W., Lauth, M.R., Pereira, D.A., Franke, A.E. (1987) The Yarrowia lipolytica LEU2 gene. *Curr. Genet.* 11:377-383.

66. de Boer, H.A., Comstock, L.J., Vasser, M. (1993) The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci.* 80:21-25.

67. De Louvencourt, L., Fukuhara, H., Heslot, H., Wesolowski, M. (1983) Transformation of *Kluyveromyces lactis* by killer plasmid DNA. *J. Bacteriol.* 154:737-742.

68. Deasy, B.M., Huard, J. (2002) Gene therapy and tissue engineering based on muscle-derived stem cells. *Curr. Opin. Mol. Ther.* 4:382-389.

69. Delahunty, C., Ankener, W., Deng, Q., Eng, J., Nickerson, D.A. (1996) Testing the feasibility of DNA typing for human identification by PCR and an oligonucleotide ligation assay. *Am. J. Human Genetics* 58:1239-1246.

70. Deutscher, M.P., Simon, M.I., Abelson, J.N., eds. (1990) Guide to Protein Purification: Methods in Enzymology. (Methods in Enzymology Series, Vol 182). Academic Press.

71. Dieffenbach, C.W., Dveksler, G.S., eds. (1995) PCR Primer: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

72. Dijkema, R., van der Meide, P.H., Pouwels, P.H., Caspers, M., Dubbeld, M., Schellekens, H. (1985) Cloning and expression of the chromosomal immune interferon gene of the rat. *EMBO J.* 4:761-767.

73. Doerfler, W., Bohm, P., eds. (1987) The Molecular Biology Of Baculoviruses. Springer-Verlag, Inc.

74. Doll, A., Grzeschik, K.H. (2001) Characterization of two novel genes, WBSCR20 and WBSCR22, deleted in Williams-Beuren syndrome. *Cytogenet. Cell Genet.* 95:20-27.

75. Doolittle, R.F., Abelson, J.N., Simon, M.I., eds. (1996) Computer Methods for Macromolecular Sequence Analysis. 1st ed. Academic Press.

76. Ducrest, A.L., Suzutorisz, H., Lingner, J., Nabholz, M. (2002) Regulation of the human telomerase reverse transcriptase gene. *Oncogene* 21:541-52.

77. Dutoit, V., Taub, R.N., Papadopoulos, K.P., Talbot, S., Keohan, M.L., Brehm, M., Gnjatic, S., Harris, P.E., Bisikirska, B., Guillaume, P., Cerottini, J.C., Hesdorffer, C.S., Old, L.J., Valmori, D. (2002) Multiepitope CD8+ T cell response to an NY-ESO-1 peptide vaccine results in imprecise tumor targeting. *J. Clin. Invest.* 110:1813-1822.

78. Egilsson, V., Gudnason, V., Jonasdottir, A., Ingvarsson, S., Andresdottir, V. (1986) Catabolite repressive effects of 5-thio-D-glucose on *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* 132:3309-3313.

79. Ehrhardt, G.R., Korherr, C., Wieler, J.S., Knaus, M., Schrader, J.W. (2001) A novel potential effector of M-Ras and p21 Ras negatively regulates p21 Ras-mediated gene induction and cell growth. *Oncogene* 20:188-197.

80. Espejo, A., Cote, J., Bednarek, A., Richard, S., Bedford, M.T. (2002) A protein-domain microarray identifies novel protein-protein interactions. *Biochem. J.* 367:697-702.

81. Everett, R.D., Meredith, M., Orr, A., Cross, A., Kathoria, M., Parkinson, J. (1997) A novel ubiquitin-specific protease is dynamically associated with the PML nuclear domain and binds to a herpesvirus regulatory protein. *EMBO J.* 16:1519-1530.

82. Fanning, A.S., Anderson, J.M. (1999) Protein modules as organizers of membrane structure. *Curr. Opin. Cell Biol.* 11:432-439.

83. Fields, S., Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340:245-246.

84. Fisch, P., Forster, A., Sherrington, P.D., Dyer, M.J., Rabbitts, T.H. (1993) The chromosomal translocation t(X;14)(q28;q11) in T-cell pro-lymphocytic leukaemia breaks within one gene and activates another. *Oncogene* 8:3271-3276.

85. Fishman, P.S., Oyler, G.A. (2002) Significance of the parkin gene and protein in understanding Parkinson's disease. *Curr. Neurol. Neurosci. Rep.* 2:296-302.

86. Forgac, M. (1999) Structure and properties of the vacuolar (H+)-ATPases. *J. Biol. Chem.* 274:12,951-12,954.

87. Frank, I. (2002) Antivirals against HIV-1. *Clin. Lab. Med.* 22:741-757.

88. Frithz, G., Ericsson, P., Ronquist, G. (1976) Serum adenylate kinase activity in the early phase of acute myocardial infarction. *Ups J Med Sci.* 81:155-158.

89. Funakoshi, I., Kato, H., Horie, K., Yano, T., Hori, Y., Kobayashi, H., Inoue, T., Suzuki, H., Fukui, S., Tsukahara, M., et al. (1992) Molecular cloning of cDNAs for human fibroblast nucleotide pyrophosphatase. *Arch. Biochem. Biophys.* 295:180-187.

90. Furth, P.A., Shamay, A., Wall, R.J., Hennighausen, L. (1992) Gene transfer into somatic tissues by jet injection. *Anal. Biochem.* 205:365-368.

91. Gaillardin, C., Ribet, A.M. (1987) LEU2 directed expression of beta-galactosidase activity and phleomycin resistance in *Yarrowia lipolytica. Curr. Genet.* 11:369-375.

92. Gao, X., Nawaz, Z. (2002) Progesterone receptors - animal models and cell signaling in breast cancer: Role of steroid receptor coactivators and corepressors of progesterone receptors in breast cancer. *Breast Cancer Res.* 4:182-186.

93. Gao, Y., Melki, R., Walden, P.D., Lewis, S.A., Ampe, C., Rommelaere, H., Vandekerckhove, J., Cowan, N.J. (1994) A novel cochaperonin that modulates the ATPase activity of cytoplasmic chaperonin. *J. Cell Biol.* 125:989-996.

94. Gaudilliere, B., Shi, Y., Bonni, A. (2002) RNA interference reveals a requirement for MEF2A in activity-dependent neuronal survival. *J. Biol. Chem.* 277:46,442-46,446.

95. Gavrieli, Y., Sherman, Y., Ben-Sasson, S.A. (1992) Identification of programmed cell death *in situ* via specific labeling of nuclear DNA fragmentation. *J. Cell Biol.* 119:493-501.

96. Geffen D.B., Man S. (2002) New drugs for the treatment of cancer, 1990-2001. *Isr. Med. Assoc. J.* 4:1124-31.

97. Gennaro, A., ed. (2000) Remington: The Science and Practice of Pharmacy. 20th ed. Lippincott, Williams, & Wilkins.

98. Ghofrani, H.A., Rose, F., Schermuly, R.T., Olschewski, H., Wiedemann, R., Kreckel, A., Weissmann, N., Ghofrani, S., Enke, B., Seeger, W., Grimminger, F. (2003) Oral sildenafil as long-term adjunct therapy to inhaled iloprost in severe pulmonary arterial hypertension. *J. Am. Coll. Cardiol.* 42:158-164.

99. Gillingham, A.K., Pfeifer, A.C., Munro, S. (2002) CASP, the alternatively spliced product of the gene encoding the CCAAT-displacement protein transcription factor, is a Golgi membrane protein related to giantin. *Mol. Biol. Cell* 13:3761-3774.

100. Gingras, M.C., Lapillonne, H., Margolin, J.F. (2002) TREM-1, MDL-1, and DAP12 expression is associated with a mature stage of myeloid development. *Mol. Immunol.* 38:817-824.

101. Girschick, H.J., Grammer, A.C., Nanki, T., Vazquez, E., Lipsky, P.E. (2002) Expression of recombination activating genes 1 and 2 in peripheral B cells of patients with systemic lupus erythematosus. *Arthritis. Rheum.* 46:1255-1263.

102. Gmeiner, W.H., Horita, D.A. (2001) Implications of SH3 domain structure and dynamics for protein regulation and drug design. *Cell Biochem. Biophys.* 35:127-140.

103. Goeddel, D.V., Heyneker, H.L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D.G., Ross, M.J., Mizzari, G., Crea, R., Seeburg, P.H. (1979) Direct expression in *E. coli* of a DNA sequence coding for human growth hormone. *Nature* 281:544-548.

104. Goldstein, L.S.B., Yang, Z. (2000) Microtubule-based transport systems in neurons: the roles of kinesins and dyneins. *Annu. Rev. Neurosci.* 23:39-71.

105. Golovkina, T.V., Chervonsky, A., Dudley, J.P., Ross, S.R. (1992) Transgenic moue mammary tumor virus superantigen expression prevents viral infection. *Cell* 69:637-645.

106. Gonnet, G.H., Cohen, M.A., Benner, S.A. (1992) Exhaustive matching of the entire protein sequence database. *Science* 256:1443-1445.

107. Gordan, J.D., Vonderheide, R.H. (2002) Universal tumor antigens as targets for immunotherapy. *Cytotherapy* 4:317-327.

108. Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I., Howard, B.H. (1982) The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eucaryotic cells by DNA-mediated transfection. *Proc. Natl. Acad. Sci.* 79:6777-6781.

109. Gray, T.A., Hernandez, L., Carey, A.H., Schaldach, M.A., Smithwick, M.J., Rus, K.M., Graves, J.A., Stewart, C.L., Nicholls, R.D. (2002) The

ancient source of a distinct gene family encoding proteins featuring RING and C(3)H zinc-finger motifs with abundant expression in developing brain and nervous system. *Genomics.* 66:76-86.

110.   Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C., Gelbart, W.M. (1999) <u>Introduction to Genetic Analysis</u>. 7th ed. W.H. Freeman.

111.   Griffiths, M., Beaumont, N., Yao, S.Y., Sundaram, M., Boumah, C.E., Davies, A., Kwong, F.Y., Coe, I., Cass, C.E., Young, J.D., Baldwin, S.A. (1997) Cloning of a human nucleoside transporter implicated in the cellular uptake of adenosine and chemotherapeutic drugs. *Nat. Med.* 3:89-93.

112.   Grosschedl, R., Baltimore, D. (1985) Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. *Cell* 41:885-897.

113.   Grosveld, F., Kollias, G., eds. (1992) <u>Transgenic Animals</u>. 1st ed. Academic Press.

114.   Gustin, K., Burk, R.D. (1993) A rapid method for generating linker scanning mutants utilizing PCR. *Biotechniques* 14:22-24.

115.   Hacia, J.G. (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genetics* 21:42-47.

116.   Hadano, S., Yanagisawa, Y., Skaug, J., Fichter, K., Nasir, J., Martindale, D., Koop, B.F., Scherer, S.W., Nicholson, D.W., Rouleau, G.A., Ikeda, J., Hayden, M.R. (2001) Cloning and characterization of three novel genes, ALS2CR1, ALS2CR2, and ALS2CR3, in the juvenile amyotrophic lateral sclerosis (ALS2) critical region at chromosome 2q33-q34: candidate genes for ALS2. *Genomics* 71:200-213.

117.   Hall, M., Mickey, D.D., Wenger, A.S., Silverman, L.M. (1985) Adenylate kinase: an oncodevelopmental marker in an animal model for human prostatic cancer. *Clin. Chem.* 31:1689-1691.

118.   Ham, R.G., McKeehan, W.L. (1979) Media and growth requirements. *Methods Enzymol.* 58:44-93.

119.   Hanada, T., Lin, L., Tibaldi, E.V., Reinherz, E.L., Chishti, A.H. (2000) GAKIN, a novel kinesin-like protein associates with the human homologue of the Drosophila discs large tumor suppressor in T lymphocytes. *J. Biol. Chem.* 275:28,774-28,784.

120.   Harlow, E., Lane, D., eds. (1988) Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory.

121.   Harlow, E., Lane, D., Harlow, E., eds. (1998) Using Antibodies: A Laboratory Manual: Portable Protocol NO. I. Cold Spring Harbor Laboratory.

122.   Hartmann, G., Endres, S., eds. (1999) Manual of Antisense Methodology (Perspectives in Antisense Science). 1st ed. Kluwer Law International.

123.   Hassanzadeh, G.H.G., De Silva, K.S., Dambly-Chudiere, C., Brys, L., Ghysen, A., Hamers, R., Muyldermans, S., De Baetselier, P. (1998) Isolation and characterization of single-chain Fv genes encoding antibodies specific for Drosophila Poxn protein. FEBS Lett. 437:75-80.

124.   Hawes, J.W., Jaskiewicz, J., Shimomura, Y., Huang, B., Bunting, J., Harper, E.T., Harris, R.A. (1996) Primary structure and tissue-specific expression of human beta-hydroxyisobutyryl-coenzyme A hydrolase. J. Biol. Chem. 271:26,430-26,434.

125.   Heath, J.K., White, S.J., Johnstone, C.N., Catimel, B., Simpson, R.J., Moritz, R.L., Tu, G.F., Ji, H., Whitehead, R.H., Groenen, .L.C., Scott, A.M., Ritter, G., Cohen, L., Welt, S., Old, L.J., Nice, E.C., Burgess, A.W. (1997) The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. Proc. Natl. Acad. Sci. 94:469-474.

126.   Heiser, A., Coleman, D., Dannull, J., Yancey, D., Maurice, M.A., Lallas, C.D., Dahm, P., Niedzwiecki, D., Gilboa, E., Vieweg, J. (2002) Autologous dendritic cells transfected with prostate-specific antigen RNA stimulate CTL responses against metastatic prostate tumors. J. Clin. Invest. 109:409-417.

127.   Henningson, C.T. Jr., Stanislaus, M.A., Gewirtz, A.M. (2003) Embryonic and adult stem cell therapy. J. Allergy Clin. Immunol. 111:S745-S753.

128.   Hinnen, A., Hicks, J.B., Fink, G.R. (1978) Transformation of yeast. Proc. Natl. Acad. Sci. 75:1929-1933.

129.   Hirsch, D.S., Pirone, D.M., Burbelo, P.D. (2001) A new family of Cdc42 effector proteins, CEPs, function in fibroblast and epithelial cell shape changes. J. Biol. Chem. 276:875-883.

130.   Ho, L.W., Carmichael, J., Swartz, J., Wyttenbach, A., Rankin, J., Rubinsztein, D.C. (2001) The molecular biology of Huntington's disease. *Psychol. Med.* 31:3-14.

131.   Hollis, G.F., Evans, R.J., Stafford-Hollis, J.M., Korsmeyer, S.J., McKearn, J.P. (1989) Immunoglobulin lambda light-chain-related genes 14.1 and 16.1 are expressed in pre-B cells and may encode the human immunoglobulin omega light-chain protein. *Proc. Natl. Acad. Sci.* 86:5552-5556.

132.   Hong, G.F. (1982) Sequencing of large double-stranded DNA using the dideoxy sequencing technique. *Biosci. Rep.* 2:907-912.

133.   Hoogenboom, H.R., de Bruin, A.P., Hufton, S.E., Hoet, R.M., Arends, J.W., Roovers, R.C. (1998) Antibody phage display technology and its applications. *Immunotechnology* 4:1-20.

134.   Hooper, M.L. (1993) <u>Embryonal Stem Cells: Introducing Planned Changes into the Animal Germline</u>. Gordon & Breach Science Pub.

135.   Hoozemans, J.J., Veerhuis, R., Rozemuller, A.J., Eikelenboom, P. (2002) The pathological cascade of Alzheimer's disease: the role of inflammation and its therapeutic implications. *Drugs Today (Barc)* 38:429-443.

136.   Houseman, B.T., Huh, J.H., Kron, S.J., Mrksich, M. (2002) Peptide chips for the quantitative evaluation of protein kinase activity. *Nature Biotechnol.* 20:270-274.

137.   Howard, G.C., Bethell, D.R. (2000) <u>Basic Methods in Antibody Production and Characterization</u>. CRC Press.

138.   Huynh, D.P., Yang, H.T., Vakharia, H., Nguyen, D., Pulst, S.M. (2003) Expansion of the polyQ repeat in ataxin-2 alters its Golgi localization, disrupts the Golgi complex and causes cell death. *Hum. Mol. Genet.* 12:1485-1496.

139.   Ikeda, A., Nishina, P.M., Naggert, J.K. (2002) The tubby-like proteins, a family with roles in neuronal development and function. *J. Cell Sci.* 115(Pt 1):9-14.

140.   Ito, H., Fukuda, Y., Murata, K., Kimura, A. (1978) Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* 153:163-168.

141.   Jameson, D.M., Sawyer, W.H. (1995) Fluorescence anisotropy applied to biomolecular interactions. *Methods Enzymol.* 246:283-300.

142.   Janeway, C.A., Travers, P. Walport, M. Shlomchik, M. (2001) Immunobiology. 5th ed. Garland Publishing.

143.   Jeffery, P., Zhu, J. (2002) Mucin-producing elements and inflammatory cells. *Novartis Found. Symp.* 248:51-75, 277-82.

144.   Jimbo, T., Kawasaki, Y., Koyama, R., Sato, R., Takada, S., Haraguchi, K., Akiyama, T. (2002) Identification of a link between the tumour suppressor APC and the kinesin superfamily. *Nat. Cell Biol.* 4:323-327.

145.   Joberty, G., Perlungher, R.R., Macara, I.G. (1999) The Borgs, a new family of Cdc42 and TC10 GTPase-interacting proteins. *Mol. Cell Biol.* 19:6585-6597.

146.   Johns, T.G., Bernard, C.C. (1997) Binding of complement component Clq to myelin oligodendrocyte glycoprotein: a novel mechanism for regulating CNS inflammation. *Mol. Immunol.* 34:33-38.

147.   Jolliffe, C.N., Harvey, K.F., Haines, B.P., Parasivam, G., Kumar, S. (2000) Identification of multiple proteins expressed in murine embryos as binding partners for the WW_domains of the ubiquitin-protein ligase Nedd4. *Biochem. J.* 351:557-565.

148.   Jones, D.H., Winistorfer, S.C. (1992) Recombinant circle PCR and recombination PCR for site-specific mutagenesis without PCR product purification. *Biotechniques* 12:528-530.

149.   Jones, P., ed. (1998a) Vectors: Cloning Applications: Essential Techniques, John Wiley & Son, Ltd.

150.   Jones, P., ed. (1998b) Vectors: Expression Systems: Essential Techniques, John Wiley & Son, Ltd.

151.   Jost, C.R., Kurucz I., Jacobus, C.M., Titus, J.A., George, A.J., Segal, D.M. (1994) Mammalian expression and secretion of functional single-chain Fv molecules. *J. Biol. Chem.* 269:26,267-26,273.

152.   Joulin, V., Richard-Foy, H. (1995) A new approach to isolate genomic control regions. Application to the GATA transcription factor family. *Eur. J. Biochem.* 232:620-626.

153.   Jurcic, J.G., Cathcart, K., Pinilla-Ibarz, J., Scheinberg, D.A. (2000) Advances in immunotherapy of hematlogic malignancies: cellular and humoral approaches. *Curr. Opin. Hematol.* 7:247-254.

154.   Jury, J.A., Perry, A.C., Hall, L. (1999) Identification, sequence analysis and expression of transcripts encoding a putative metalloproteinase, eMDC II, in human and macaque epididymis. *Mol. Hum. Reprod.* 5:1127-1134.

155.   Kabat, E.A., Wu T.T. (1991) Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. *J. Immunol.* 147:1709-1719.

156.   Kamitani, T., Nguyen, H.P., Yeh, E.T. (1997) Preferential modification of nuclear proteins by a novel ubiquitin-like molecule. *J. Biol. Chem.* 272:14,001-14,004.

157.   Kantoff, P.W., Halabi, S., Farmer, D.A., Hayes, D.F., Vogelzang, N.A., Small, E.J. (2001) Prognostic significance of reverse transcriptase polymerase chain reaction for prostate-specific antigen in men with hormone-refractory prostate cancer. *J. Clin. Oncol.* 9:3025-3028.

158.   Kao, P.N., Chen, L., Brock, G., Ng, J., Kenny, J., Smith, A.J., Corthesy, B. (1994) Cloning and expression of cyclosporin A- and FK506-sensitive nuclear factor of activated T-cells: NF45 and NF90. *J. Biol. Chem.* 269:20,691-20,699.

159.   Karanazanashvili, G., Abrahamsson, P. (2003) Prostate specific antigen and human glandular kallikrein 2 in early detection of prostate cancer. *J. Urol.* 169:445-457.

160.   Kari, C., Chan, T.O., Rocha de Quadros, M., Rodeck, U. (2003) Targeting the epidermal growth factor receptor in cancer: apoptosis takes center stage. *Cancer Res.* 63:1-5.

161.   Kelly, J.M., Hynes, M.J. (1985) Transformation of *Aspergillus niger* by the mdS gene of *Aspergillus nidulans. EMBO J.* 4:475-479.

162.   Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T.J., Tanaka, T., Page, D.C. (1998) A map of 75 human ribosomal protein genes. *Genome Res.* 8:509-523.

163.   Keown, W.A., Campbell, C.R., Kucherlapati, R.S. (1990) Methods for introducing DNA into mammalian cells. *Methods Enzymol.* 185:527-537.

164.   Kibbe, A.H., ed. (2000) <u>Handbook of Pharmaceutical Excipients.</u> 3rd ed. Pharmaceutical Press.

165.   Kirkpatrick, K.L., Mokbel, K. (2001) The significance of human telomerase reverse transcriptase (hTERT) in cancer. *Eur. J. Surg. Oncol.* 27:754-760.

166.   Kirsch, K.H., Georgescu, M.M., Ishimaru, S., Hanafusa, H. (1999) CMS: an adapter molecule involved in cytoskeletal rearrangements. *Proc. Natl. Acad. Sci.* 96:6211-6216.

167.   Kiryu-Seo, S., Sasaki, M., Yokohama, H., Nakagomi, S., Hirayama, T., Aoki, S., Wada, K., Kiyama, H. (2000) Damage-induced neuronal endopeptidase (DINE) is a unique metallopeptidase expressed in response to neuronal damage and activates superoxide scavengers. *Proc. Natl. Acad. Sci.* 97:4345-4350.

168.   Klarman, G.J., Hawkins, M.E., Le Grice, S.F. (2002) Uncovering the complexities of retroviral ribonuclese H reveals its potential as a therapeutic target. *AIDS Rev.* 4:183-194.

169.   Knutson, K.L., Schiffman, K., Disis, M.L. (2001) Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. *J. Clin. Invest.* 107:477-484.

170.   Kobayashi, M., Takezawa, S., Hara, K., Yu, R.T., Umesono, Y., Agata, K., Taniwaki, M., Yasuda, K., Umesono, K. (1999) Identification of a photoreceptor cell-specific nuclear receptor. *Proc. Natl. Acad. Sci.* 96:4814-4819.

171.   Kolonin, M.G., Finley, R.L. Jr. (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc. Natl. Acad. Sci.* 95:14,266-14,271.

172.   Korner, C., Knauer, R., Stephani, U., Marquardt, T., Lehle, L., von Figura, K. (1999) Carbohydrate deficient glycoprotein syndrome type IV: deficiency of dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase. *EMBO J.* 18:6816-6822.

173.   Kothapalli, R., Buyuksal, I., Wu, S.Q., Chegini, N., Tabibzadeh, S. (1997) Detection of ebaf, a novel human gene of the transforming growth

factor beta superfamily association of gene expression with endometrial bleeding. *J. Clin. Invest.* 99:2342-2350.

174.　　　Kovalenko, O.V., Golub, E.I., Bray-Ward, P., Ward, D.C., Radding, C.M. (1997) A novel nucleic acid-binding protein that interacts with human rad51 recombinase. *Nucleic Acids Res.* 25:4946-4953.

175.　　　Kratzschmar, J., Lum, L., Blobel, C.P. (1996) Metargidin, a membrane-anchored metalloprotease-disintegrin protein with an RGD integrin binding sequence. *J. Biol. Chem.* 271:4593-4596.

176.　　　Ku, D.H., Kagan, J., Chen, S.T., Chang, C.D., Baserga, R., Wurzel, J. (1990) The human fibroblast adenine nucleotide translocator gene. Molecular cloning and sequence. *J. Biol. Chem.* 265:16,060-16,063.

177.　　　Kuisle, O., Quiñoá, E., Rigura, R. (1999) Solid phase synthesis of depsides and depsipeptides. *Tetrahedron Lett.* 40:1203-1206.

178.　　　Kunze, G. et al., (1985) Transformation of the industrially important yeasts *Candida* maltosa and *Pichia* guilliermondii. *J. Basic Microbiol.* 25:141-144.

179.　　　Kurtz, M.B., Cortelyou, M.W., Kirsch, D.R. (1986) Integrative transformation of Candida albicans, using a cloned *Candida* ADE2 gene. *Mol. Cell. Biol.* 6:142-149.

180.　　　Kyo, S., Takakura, M., Inoue, M. (2000) Telomerase activity in cancer as a diagnostic and therapeutic target. *Histol. Histopathol.* 15:813-824.

181.　　　Lander, E.S. (1999) Array of hope. *Nature Genetics* 21:3-4.

182.　　　Lander, E.S., Linton, L..M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R.., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S.., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M.,

Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., Szustakowki, J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J.; International

Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome *Nature* 409:860-921.

183.   Lasham, A., Moloney, S., Hale, T., Homer, C., Zhang, Y.F., Murison, J.G., Braithwaite, A.W., Watson, J. (2003) The Y-box binding protein YB1: A potential negative regulator of the p53 tumor suppressor. *J. Biol. Chem.* Epub ahead of print, June 30, 2003.

184.   Lashkari, A., Smith, A.K., Graham, J.M. Jr. (1999) Williams-Beuren syndrome: an update and review for the primary physician. *Clin. Pediatr.* 38:189-208.

185.   Lavedan, C. (1998) The synuclein family. *Genome Res.* 8:871-880.

186.   Lebacq-Verheyden, A.M., Kasprzyk, P.G., Raum, M.G., Van Wyke Coelingh, K., Lebacq, J.A., Battey, J.F. (1988) Posttranslational processing of endogenous and of baculovirus-expressed human gastrin-releasing peptide precursor. *Mol. Cell. Biol.* 8:3129-3135.

187.   Lees-Miller, S.P., Anderson, C.W. (1989) Two human 90-kDa heat-shock proteins are phosphorylated in vivo at conserved serines that are phosphorylated *in vitro* by casein kinase II. *J. Biol. Chem.* 264:2431-2437.

188.   Lerch, M.M., Gorelick, F.S. (2000) Early trypsinogen activation in acute pancreatitis. *Med. Clin. North Amer.* 84:549-563.

189.   Lerner, R.A. (1982) Tapping the immunological repertoire to produce antibodies of predetermined specificity. *Nature* 299:592-596.

190.   Li, E., Bestagno, M., Burrone, O. (1996) Molecular cloning and characterization of a transmembrane surface antigen in human cells. *Eur. J. Biochem.* 238:631-638.

191.   Lim, D., Orlova, M., Goff, S.P. (Aug. 2002) Mutations of the RNase H C helix of the Moloney murine leukemia virus reverse transcriptase reveal defects in polypurine tract recognition. *J. Virol.* 76:8360-8373.

192.   Lin, B., Rommens, J.M., Graham, R.K., Kalchman, M., MacDonald, H., Nasir, J., Delaney, A., Goldberg, Y.P., Hayden, M.R. (1993) Differential 3' polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression. *Hum. Mol. Genet.* 2:1541-1545.

193.   Lin, W.J., Gary, J.D., Yang, M.C., Clarke, S., Herschman, H.R. (1996) The mammalian immediate-early TIS21 protein and the leukemia-

associated BTG1 protein interact with a protein-arginine N-methyltransferase. *J. Biol. Chem.* 271:15,034-15,044.

194.   Lin, X., Sikkink, R.A., Rusnak, F., Barber, D.L. (1999) Inhibition of calcineurin phosphatase activity by a calcineurin B homologous protein. *J. Biol. Chem.* 274:36,125-36,131.

195.   Linnenbach, A.J., Seng, B.A., Wu, S., Robbins, S., Scollon, M., Pyrc, J.J., Druck, T., Huebner, K. (1993) Retroposition in a family of carcinoma-associated antigen genes. *Mol. Cell Biol.* 13:1507-1515.

196.   Linstedt, A.D., Hauri, H.P. (1993) Giantin, a novel conserved Golgi membrane protein containing a cytoplasmic domain of at least 350 kDa. *Mol. Biol. Cell* 4:679-693.

197.   Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R., Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nature Genetics* 21:20-24.

198.   Liu A.Y., Robinson R.R., Hellstrom K.E., Murray E.D. Jr., Chang C.P., Hellstrom I. (1987a) Chimeric mouse-human IgG1 antibody that can mediate lysis of cancer cells. *Proc. Natl. Acad. Sci.* 84:3439-3443.

199.   Liu, A.Y., Robinson, R.R., Murray, E.D. Jr., Ledbetter, J.A., Hellstrom, I., Hellstorm, K.E. (1987b) Production of a mouse-human chimeric monoclonal antibody to CD20 with potent Fc-dependent biologic activity. *J. Immunol.* 139:3521-3526.

200.   Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darness, J. (1999) <u>Molecular Cell Biology</u>. 4th ed. W H Freeman & Co.

201.   Loeffen, J.L., Triepels, R.H., van den Heuvel, L.P., Schuelke, M., Buskens, C.A., Smeets, R.J., Trijbels, J.M., Smeitink, J.A. (1998) cDNA of eight nuclear encoded subunits of NADH:ubiquinone oxidoreductase: human complex I cDNA characterization completed. *Biochem. Biophys. Res. Commun.* 253:415-422.

202.   Los, M., Burek, C.J., Stroh, C., Benedyk, K., Hug, H., Mackiewicz. (2003) Anticancer drugs of tomorrow: apoptotic pathways as targets for drug design. *Drug Discov. Today* 15:67-77.

203.   Lovering R, Trowsdale J. (1991) A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell lines. *Nucleic Acids Res.* 19:2921-2928.

204.   Luckow, V., Summers, M. (1988) Trends in the development of baculovirus expression vectors. *Bio/Technology* 6:47-55.

205.   MacBeath, G., Schreiber. S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289:1760-1763.

206.   Machesky, L.M., Reeves, E., Wientjes, F., Mattheyse, F.J., Grogan, A., Totty, N.F., Burlingame, A.L., Hsuan, J.J., Segal, A.W. (1999) Mammalian actin-related protein 2/3 complex localizes to regions of lamellipodial protrusion and is composed of evolutionarily conserved proteins. *Biochem. J.* 328:105-112.

207.   Machiels, J.P., van Baren, N., Marchand, M. (2002) Peptide-based cancer vaccines. *Semin. Oncol.* 29:494-502.

208.   Mackay, A., Jones, C., Dexter, T., Silva, R.L., Bulmer, K., Jones, A., Simpson, P., Harris, R.A., Jat, P.S., Neville, A.M., Reis, L.F., Lakhani, S.R., O'Hare, M.J. (2003) cDNA microarray analysis of genes associated with ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. *Oncogene* 22:2680-2688.

209.   Maeda, S., Kawai, T., Obinata, M., Fujiwara, H., Horiuchi, T., Saeki, Y., Sato, Y., Furusawa, M. (1985) Production of human alpha-interferon in silkworm using a baculovirus vector. *Nature* 315:592-594.

210.   Mahajan, M.A., Murray, A., Samuels, H.H. (2002) NRC-interacting factor 1 is a novel cotransducer that interacts with and regulates the activity of the nuclear hormone receptor coactivator NRC. *Mol. Cell Biol.* 22:6883-6894.

211.   Mahimkar, R.M., Baricos, W.H., Visaya, O., Pollock, A.S., Lovett, D.H. (2000) Identification, cellular distribution and potential function of the metalloprotease-disintegrin MDC9 in the kidney. *J. Am. Soc. Nephrol.*, 11:595-603.

212.   Mahnensmith, R.L., Aronson, P.S. (1985) Interrelationships among quinidine, amiloride, and lithium as inhibitors of the renal Na+-H+ exchanger. *J. Biol. Chem.* 260:12,586-12,592.

213.   Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science* 298:1912-1934.

214.   Marotti, K.R., Tomich, C.S. (1989) Simple and efficient oligonucleotide-directed mutagenesis using one primer and circular plasmid DNA template. (1989) *Gene Anal. Tech.* 6:67-70.

215.   Martel-Pelletier, J., Welsch, D.J., and Pelleteir, J.P. (2001) Metalloproteases and inhibitors in arthritic diseases. *Best Pract. Res. Clin. Rheumatol.* 15:805-829.

216.   Martin, B.M., Tsuji, S., LaMarca, M.E., Maysak, K., Eliason, W., Ginns, E.I. (1988) Glycosylation and processing of high levels of active human glucocerebrosidase in invertebrate cells using a baculovirus expression vector. *DNA* 7:99-106.

217.   Massari, M.E., Rivera, R.R., Voland, J.R., Quong, M.W., Breit, T.M., van Dongen, J.J,, de Smit, O., Murre, C. (1998) Characterization of ABF-1, a novel basic helix-loop-helix transcription factor expressed in activated B lymphocytes. *Mol. Cell Biol.* 18:3130-3139.

218.   Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., Lukyanov, S.A. (1999) Fluorescent proteins from nonbioluminescent *Anthozoa* species. *Nat. Biotechnol.* 17:969-973.

219.   Mayer, B.J. (2001) SH3 domains: complexity in moderation. *J. Cell Sci.* 114:1253-1263.

220.   Mayer, T.U., Kapoor, T.M., Haggarty, S.J., King, R.W., Schreiber, S.L., Mitchison, T.J. (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* 286:971-974.

221.   McGraw, R.A. III (1984) Dideoxy DNA sequencing with end-labeled oligonucleotide primers. *Anal. Biochem.* 143:298-303.

222.   McKusick, V.A.. (2003) OMIM: Online Mendelian Inheritance in Man http:www.ncbi.nlm.nih.gov, #104300.

223.   McPherson, M.J., Møller, S.G., Benyon, R., Howe, C. (2000) <u>PCR Basics: From Background to Bench</u>. Springer Verlag.

224.   Merla, G., Ucla, C., Guipponi, M., Reymond, A. (2002) Identification of additional transcripts in the Williams-Beuren syndrome critical region. *Hum. Genet.* 110:429-438.

225.   Miki, H., Setou, M., Kaneshiro, K., Hirokawa, N. (2001) All kinesin superfamily protein, KIF, genes in mouse and human. *Proc. Natl. Acad. Sci.* 98:7004-7011.

226.   Milam, A.H., Rose, L., Cideciyan, A.V., Barakat, M.R., Tang, W.X.,
Gupta, N., Aleman, T.S., Wright, A.F., Stone, E.M., Sheffield, V.C., Jacobson,
S.G. (2002) The nuclear receptor NR2E3 plays a role in human retinal
photoreceptor differentiation and degeneration. *Proc. Natl. Acad. Sci.* 99:473-
478.

227.   Milligan, J.F., Matteucci, M.D., Martin, J.C. (1993) Current concepts
in antisense drug design. *J. Med. Chem.* 36:1923-1937.

228.   Mitch, W.E., Goldberg, A.L. (1996) Mechanisms of muscle wasting.
The role of the ubiquitin-proteasome pathway. *N. Engl. J. Med.* 335:1897-
1905.

229.   Mitchell, D.A., Nair, S.K. (2000) RNA-transfected dendritic cells in
cancer immunotherapy. *J. Clin. Invest.* 106:1065-1069.

230.   Miyajima A. (2002) Functional analysis of yeast homologue gene
associated with human DNA helicase causative syndromes. *Kokuritsu
Iyakuhin Shokuhin Eisei Kenkyusho Hokoku* 120:53-74.

231.   Miyajima, A., Schreurs, J., Otsu, K., Kondo, A., Arai, K., Maeda, S.
(1987) Use of the silkworm, *Bombyx mori*, and an insect baculovirus vector
for high-level expression and secretion of biologically active mouse
interleukin-3. *Gene* 58:273-281.

232.   Monfardini, C., Schiavon, O., Caliceti, P., Morpurgo, M., Harris, J.M.,
Veronese, F.M. (1995) A branched monomethoxypoly(ethylene glycol) for
protein modification. *Bioconjugate Chem.* 6:62-69.

233.   Mori, N. (1997) Neuronal growth-associated proteins in neural
plasticity and brain aging. *Nihon Shinkei Seishin Yakurigaku Zasshi* 17:159-
167.

234.   Mortlock, D.P., Nelson, M.R., Innis, J.W. (1996) An efficient method
for isolating putative promoters and 5' transcribed sequences from large
genomic clones. *Genome Res.* 6:327-335.

235.   Murphy, D., Carter, D.A., eds. (1993) <u>Transgenesis Techniques:
Principles and Protocols</u>. Humana Press.

236.   Myers, E.W., Miller, W. (1988) Optimal alignments in linear space.
Comput. Appl. Biosci. 4:11-7.

237.   Nagata, K., Kawase, H., Handa, H., Yano, K., Yamasaki, M., Ishimi,
Y., Okuda, A., Kikuchi, A., Matsumoto, K. (1995) Replication factor encoded

by a putative oncogene, set, associated with myeloid leukemogenesis. *Proc. Natl. Acad. Sci.* 92:4279-4283.

238.　　　Naora, H. (1999) Involvement of ribosomal proteins in regulating cell growth and apoptosis: translational modulation or recruitment for extraribosomal activity? *Immunol. Cell Biol.* 77:197-205.

239.　　　Needleman, S.B., Wunch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48:443-453.

240.　　　Nelson, N., Harvey, W.R. (1999) Vacuolar and plasma membrane proton-adenosine triphosphatases. *Physiol. Rev.* 79:361-385.

241.　　　Nishiyama, H., Higashitsuji, H., Yokoi, H., Itoh, K., Danno, S., Matsuda, T., Fujita, J. (1997) Cloning and characterization of human CIRP (cold-inducible RNA-binding protein) cDNA and chromosomal assignment of the gene. *Gene* 204:115-120.

242.　　　Noma, T., Fujisawa, K., Yamashiro, Y., Shinohara, M., Nakazawa, A., Gondo, T., Ishihara, T., Yoshinobu, K. (2001) Structure and expression of human mitochondrial adenylate kinase targeted to the mitochondrial matrix. *Biochem. J.* 358:225-232.

243.　　　Notredame, C., Higgins, D., Heringa, J. (2000) T-Coffee: A novel method for multiple sequence alignments. *J. Molec. Biol.* 302:205-217.

244.　　　Okayama, H., Berg, P. (1983) A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. *Mol. Cell. Biol.* 3:280-289.

245.　　　Oksenberg, J.R., Barcellos, L.F., Hauser, S.L. (1999) Genetic aspects of multiple sclerosis. *Semin. Neurol.* 19:281-288.

246.　　　Oliver, C.J., Shenolikar, S. (1998) Physiologic importance of protein phosphatase inhibitors. *Frontiers in Bioscience* 3:961-972.

247.　　　O'Neil, N.J., Martin, R.L., Tomlinson, M.L., Jones, M.R., Coulson, A., Kuwabara, P.E. (2001) RNA-mediated interference as a tool for identifying drug targets. *Am. J. Pharmacogenomics* 1:45-53.

248.　　　O'Neill, L.A. (2002) Signal transduction pathways activated by the IL-1 receptor/toll-like receptor superfamily. *Curr. Top. Microbiol. Immunol.* 270:47-61.

249.　　　Page, D.C., Silber, S., Brown, L.G. (1999) Men with infertility caused by AZFc deletion can produce sons by intracytoplasmic sperm injection, but are likely to transmit the deletion and infertility. *Hum. Reprod.* 14:1722-1726.

250.　　　Pan, C.X., Koeneman, K.S. (1999) A novel tumor-specific gene therapy for bladder cancer. *Med. Hypothesis* 53:130-135.

251.　　　Pang, T., Wakabayashi, S., Shigekawa, M. (2001) Calcineurin homologous protein as an essential cofactor for Na+/H+ exchangers. *J. Biol. Chem* 276:17,367-17,372.

252.　　　Pang, T., Wakabayashi, S., Shigekawa, M. (2002) Expression of calcineurin B homologous protein 2 protects serum deprivation-induced cell death by serum-independent activation of Na+/H+ exchanger. *J. Biol. Chem.* 277:43,771-43,777.

253.　　　Papagerakis, S., Shabana, A.H., Depondt, J., Gehanno, P., Forest, N. (2003) Immunohistochemical localization of plakophilins (PKP1, PKP2, PKP3, and p0071) in primary oropharyngeal tumors: correlation with clinical parameters. *Hum. Pathol.* 34:565-572.

254.　　　Pearson, W.R. (2000) Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.* 132:185-219.

255.　　　Peattie, D.A., Harding, M.W., Fleming, M.A., DeCenzo, M.T., Lippke, J.A., Livingston, D.J., Benasutti, M. (1992) Expression and characterization of human FKBP52, an immunophilin that associates with the 90-kDa heat-shock protein and is a component of steroid receptor complexes. *Proc. Natl. Acad. Sci.* 89:10,974-10,978.

256.　　　Peelle, B., Gururaja, T.L., Payan, D.G., Anderson, D.C. (2001) Characterization and use of green fluorescent proteins from *Renilla mulleri* and *Ptilosarcus guernyi* for the human cell display of functional peptides. *J. Protein Chem.* 20:507-519.

257.　　　Pepin, K., Momose, F., Ishida, N., Nagata, K. (2001) Molecular cloning of horse Hsp90 cDNA and its comparative analysis with other vertebrate Hsp90 sequences. *J. Vet. Med. Sci.* 63:115-124.

258.　　　Perez Calvo, J.I., Inigo Gil, P., Giraldo Castellano, P., Torralba Cabeza, M.A., Civeira, F., Lario Garcia, S., Pocovi, M., Lara Garcia, S. (2000) Transforming growth factor beta (TGF-beta) in Gaucher's disease.

Preliminary results in a group of patients and their carrier and non-carrier relatives *Med. Clin. (Barc)* 115:601-604.

259.   Perron, H., Garson, J.A., Bedin, F., Beseme, F., Paranhos-Baccala, G., Komurian-Pradel, F., Mallet, F., Tuke, P.W., Voisset, C., Blond, J.L., Lalande, B., Seigneurin, J.M., Mandrand, B., The Collaborative Research Group on Multiple Sclerosis (1997) Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. *Proc. Natl. Acad. Sci.* 94:7583-7588.

260.   Perry, A.C., Jones, R., Hall, L. (1995) Analysis of transcripts encoding novel members of the mammalian metalloprotease-like, disintegrin-like, cysteine-rich (MDC) protein family and their expression in reproductive and non-reproductive monkey tissues. *Biochem. J.* 312(Pt 1):239-244.

261.   Pertl, U., Wodrich, H., Ruelmann, J.M., Gillies, S.D., Lode, H.N., Reisfeld, R.A. (2003) Immunotherapy with a posttranscriptionally modified DNA vaccine induces complete protection against metastatic neuroblastoma. *Blood* 101:649-654.

262.   Pfutzer, R.H., Whitcomb, D.C. (2001) SPINK1 mutations are associated with multiple phenotypes. *Pancreatology* 1:457-460.

263.   Phillips, M.I., ed. (1999a) <u>Antisense Technology, Part A. Methods in Enzymology Vol. 313</u>. Academic Press, Inc.

264.   Phillips, M.I., ed. (1999b) <u>Antisense Technology, Part B. Methods in Enzymology Vol. 314</u>. Academic Press, Inc.

265.   Pietu, G., Alibert, O., Guichard, V., Lamy, B., Bois, F., Mariage-Sampson, R., Hougatte, R., Soularue, P., Auffray, C. (1996) Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.* 6:492-503.

266.   Pinkert, C.A., ed. (1994) <u>Transgenic Animal Technology: A Laboratory Handbook</u>. Academic Press.

267.   Pisegna, J.R., Wank, S.A. (1996) Cloning and characterization of the signal transduction of four splice variants of the human pituitary adenylate cyclase activating polypeptide receptor. Evidence for dual coupling to adenylate cyclase and phospholipase C. *J. Biol. Chem.* 271:17,267-17,274.

268.   Prentki, P., Krisch, H.M. (1984) *In vitro* insertional mutagenesis with a selectable DNA fragment. *Gene* 29:303-313.

269. Price, N.T., Hall, L., Proud, C.G. (1993) Cloning of cDNA for the beta-subunit of rabbit translation initiation factor-2 using PCR. *Biochim. Biophys. Acta* 1216:170-172.

270. Qin, J., Li., L. (2003) Molecular anatomy of the DNA damage and replication checkpoints. *Radiat. Res.* 159:139-148.

271. Racevskis, J., Dill, A., Stockert, R., Fineberg, S.A. (1996) Cloning of a novel nucleolar guanosine 5'-triphosphate binding protein autoantigen from a breast tumor. *Cell. Growth Differ.* 7:271-280.

272. Ramalho-Santos, M. (2002) "Stemness" *Science* 298:597-600.

273. Raval, P. (1994) Qualitative and quantitative determination of mRNA. *J. Pharmacol. Toxicol. Methods* 32:125-127.

274. Rebbe, N.F., Ware, J., Bertina, R.M., Modrich, P., Stafford, D.W. (1987) Nucleotide sequence of a cDNA for a member of the human 90-kDa heat-shock protein family. *Gene* 53:235-245.

275. Rechid, R., Vingron, M., Argos, P. (1989) A new interactive protein sequence alignment program and comparison of its results with widely used algorithms. *Comput. Appl. Biosci.* 5:107-113.

276. Rehli, M., Krause, S.W., Kreutz, M., Andreesen, R. (1995) Carboxypeptidase M is identical to the MAX.1 antigen and its expression is associated with monocyte to macrophage differentiation. *J. Biol. Chem.* 270:15644-15649.

277. Remington, J.P. (1985) <u>Remington's Pharmaceutical Sciences</u>. 17th ed. Mack Publishing Co.

278. Ribardo, D.A., Peterson, J.W., Chopra, A.K. (2002) Phospholipase A2-activating protein--an important regulatory molecule in modulating cyclooxygenase-2 and tumor necrosis factor production during inflammation. *Indian J. Exp. Biol.* 40:129-138.

279. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C., Markham, A.F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nuc. Acids Res.* 18:2887-2890.

280. Ritter, R.C., Brenner, L.A., Tamura, C.S. (1994) Endogenous CCK and the peripheral neural substrates of intestinal satiety. *Ann. N. Y. Acad. Sci.* 713:255-267.

281.   Robertson, H.M. (1996) Members of the pogo superfamily of DNA-mediated transposons in the human genome. *Mol. Gen. Genet.* 252:761-766.

282.   Robertson, H.M., Zumpano, K.L. (1997) Molecular evolution of an ancient mariner transposon, Hsmar1, in the human genome. *Gene* 205:203-217.

283.   Roepman, R., Bernoud-Hubac, N., Schick, D.E., Maugeri, A., Berger, W., Ropers, H.H., Cremers, F.P., Ferreira, P.A. (2000) The retinitis pigmentosa GTPase regulator (RPGR) interacts with novel transport-like proteins in the outer segments of rod photoreceptors. *Hum. Mol. Genet.* 9:2095-2105.

284.   Roessler, B.J., Nosal, J.M., Smith, P.R., Heidler, S.A., Palella, T.D., Switzer, R.L., Becker, M.A. (1993) Human X-linked phosphoribosylpyrophosphate synthetase superactivity is associated with distinct point mutations in the PRPS1 gene. *J. Biol. Chem.* 268:26476-26481.

285.   Roggenkamp, R., Janowicz, Z., Stanikowski, B., Hollenberg, C.P. (1984) Biosynthesis and regulation of the peroxisomal methanol oxidase from the methylotrophic yeast *Hansenula polymorpha. Mol. Gen. Genet.* 194:489-493.

286.   Rosen, R.C., McKenna, K.E. (2002) PDE-5 inhibition and sexual response: pharmacological mechanisms and clinical outcomes. *Ann. Rev. Sex Res.* 13:36-88.

287.   Rosato, R.R., Grant, S. (2003) Histone deacetylase inhibitors in cancer therapy. *Cancer Biol. Ther.* 2:30-37.

288.   Rowland, J.M. (2002) Molecular genetic diagnosis of pediatric cancer: current and emerging methods. *Pediatr. Clin. North Am.* 49:1415-1435.

289.   Saha, S., Bardelli, A., Buckhaults, P., Velculescu, V.E., Rago, C., St Croix, B., Romans, K.E., Choti, M.A., Lengauer, C., Kinzler, K.W., Vogelstein, B. (2001) A phosphatase associated with metastasis of colorectal cancer. *Science* 294:1343-1346.

290.   Saiki, R.K, Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.

291.   Sambrook, J., Russell, D.W., Sambrook, J. (1989) <u>Molecular Cloning,</u> <u>A Laboratory Manual.</u> 2nd ed. Cold Spring Harbor Laboratory Press.

292.   Sanchez, E.R., Faber, L.E., Henzel, W.J., Pratt, W.B. (1990) The 56- 59-kilodalton protein identified in untransformed steroid receptor complexes is a unique protein that exists in cytosol in a complex with both the 70- and 90- kilodalton heat-shock proteins. *Biochemistry* 29:5145-5152.

293.   Sayers, J.R., Krekel, C., Eckstein, F. (1992) Rapid high-efficiency site-directed mutagenesis by the phosphothioate approach. *Biotechniques* 13:592-596.

294.   Schaeferling, M., Schiller, S., Paul, H., Kruschina, M., Pavlickova, M., Meerkamp, M., Giammasi, C., Kambhampati, D. (2002) Application of self- assembly techniques in the design of biocompatible protein microarray surfaces. *Electrophoresis* 23:3097-3105.

295.   Schaffer, J.E., Lodish, H.F. (1994) Expression cloning and characterization of a novel adipocyte long chain fatty acid transport protein. *Cell* 79:393-395.

296.   Schena, M., ed. (1999) <u>DNA Microarrays: A Practical Approach.</u> Oxford Univ. Press.

297.   Schena, M., ed. (2000) <u>Microarray Biochip Technology.</u> 1st ed. Eaton Publishing Co.

298.   Schlesinger, D.H. (1988a) <u>MacRomolecular Sequencing and</u> <u>Synthesis: Selected Methods and Applications.</u> Wiley-Liss.

299.   Schlesinger, D.H., ed. (1988b) <u>Current Methods in Sequence</u> <u>Comparison and Analysis, Macromolecule Sequencing and Synthesis,</u> <u>Selected Methods and Applications,</u> pp. 127-149, Alan R. Liss, Inc.

300.   Schonthal, A.H. (2001) Role of serine/threonine protein phosphatase 2A in cancer. *Cancer Lett.* 170:1-13.

301.   Seelig, H.P., Schranz, P., Schroter, H., Wiemann, C., Renz, M. (1994) Macrogolgin--a new 376 kD Golgi complex outer membrane protein as target of antibodies in patients with rheumatic diseases and HIV infections. *J. Autoimmun.* 7:67-91.

302.   Selkoe, D.J. (2001) Presenilin, Notch, and the genesis and treatment of Alzheimer's disease. *Proc. Natl. Acad. Sci.* 98:11,039-11,041.

303.   Setlow, J., Hollaender, A., eds. (1986) Genetic Engineering: Principles and Methods. Plenum Pub. Corp.

304.   Shamay, M., Barak, O., Doitsh, G., Ben-Dor, I., Shaul, Y. (2002) Hepatitis B virus pX interacts with HBXAP, a PHD finger protein to coactivate transcription. *J. Biol. Chem.* 277:9982-9988.

305.   Shao, H., Andres, D.A. (2000) A novel RalGEF-like protein, RGL3, as a candidate effector for rit and Ras. *J. Biol. Chem.* 275:26,914-26,924.

306.   Sheppard, P., Kindsvogel, W., Xu, W., Henderson, K., Schlutsmeyer, S., Whitmore, T.E., Kuestner, R., Garrigues, U., Birks, C., Roraback, J., Ostrander, C., Dong, D., Shin, J., Presnell, S., Fox, B., Haldeman, B., Cooper, E., Taft, D., Gilbert, T., Grant, F.J., Tackett, M., Krivan, W., McKnight, G., Clegg, C., Foster, D., Klucher, K.M. (2003) IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nat. Immunol.* 4:63-68.

307.   Shinnick, T.M., Sutcliffe, J.G., Green, N., Lerner, R.A. (1983) Synthetic peptide immunogens as vaccines. *Ann. Rev. Microbiol.* 37:425-446.

308.   Shorter, J., Beard, M.B., Seemann, J., Dirac-Svejstrup, A.B., Warren, G. (2002) Sequential tethering of Golgins and catalysis of SNAREpin assembly by the vesicle-tethering protein p115. *J. Cell Biol.* 157:45-62.

309.   Siebenlist, U., Simpson, R.B., Gilbert, W. (1980) *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell* 20:269-281.

310.   Siegal, G.J., Agranoff, B.W., Albers, R.W., Fisher, S.K., Uhler, M.D., eds. (1999) Basic Neurochemistry, Molecular, Cellular, and Medical Aspects. 6th ed. Lippencott, Williams & Wilkins.

311.   Sladek, R., Bader, J.A., Giguere, V. (1997) The orphan nuclear receptor estrogen-related receptor alpha is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene. *Mol. Cell Biol.* 17:5400-5409.

312.   Slavin, S., Or, R., Aker, M., Shapira, M.Y., Panigrahi, S., Symeonidis, A., Cividalli, G., Nagler, A. (2001) Nonmyeloablative stem cell transplantation for the treatment of cancer and life-threatening nonmalignant disorders: past accomplishments and future goals. *Cancer Chemother. Pharmacol.* 48:S79-S84.

313.   Smit, A.F., Riggs, A.D. (1996) Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci.* 93:1443-1448.

314.   Smith, G.E., Ju, G., Ericson, B.L., Moschera, J., Lahm, H.W., Chizzonite, R., Summers, M.D. (1985) Modification and secretion of human interleukin 2 produced in insect cells by a baculovirus expression vector. *Proc. Natl. Acad. Sci.* 82:8404-8408.

315.   Smith, T.F., Waterman, M.S. (1981) Comparison of biosequences. *Adv. Appl. Math.* 2:482-489.

316.   Soares, M.B. (1997) Identification and cloning of differentially expressed genes. *Curr. Opin. Biotechnol.* 8:542-546.

317.   Soejima, H., Kawamoto, S., Akai, J., Miyoshi, O., Arai, Y., Morohka, T., Matsuo, S., Niikawa, N., Kimura, A., Okubo, K., Mukai, T. (2001) Isolation of novel heart-specific genes using the BodyMap database. *Genomics.* 74:115-120.

318.   Soulier, S., Vilotte, J.L., L'Huillier, P.J., Mercier, J.C. (1996) Developmental regulation of murine integrin beta 1 subunit- and Hsc73-encoding genes in mammary gland: sequence of a new mouse Hsc73 cDNA. *Gene* 172:285-289.

319.   Southern, E., Mir, K., Shchepinov, M. (1999) Molecular interactions on microarrays. *Nature Genetics* 21:5-9.

320.   Stein, C.A., Kreig, A.M., eds. (1998) <u>Applied Antisense Oligonucleotide Technology</u>. Wiley-Liss.

321.   Steinhaur, C., Wingren, C., Hager, A.C., Borrebaeck, C.A. (2002) Single framework recombinant antibody fragments designed for protein chip applications. *Biotechniques, Supp.*:38-45.

322.   Stetler-Stevenson, W.G., Liotta, L.A., Kleiner, D.E. Jr. (1993) Extracellular matrix 6: role of matrix metalloproteinases in tumor invasion and metastasis. *FASEB J.* 7:1434-1441.

323.   Stewart, Z.A., Westfall, M.D., Pietenpol, J.A. (2003) Cell-cycle dysregulation and anticancer therapy. *Trends Pharmacol. Sci.* 24:139-145.

324.   Stolz, L.E., Tuan, R.S. (1996) Hybridization of biotinylated oligo(dT) for eukaryotic mRNA quantitation. *Mol. Biotechnol.* 6:225-230.

325.   Sturm, A., Dignass, A.U. (2002) Modulation of gastrointestinal wound repair and inflammation by phospholipids. *Biochim. Biophys. Acta* 1582:282-288.

326.   Stutz, F., Bachi, A., Doerks, T., Braun, I.C., Seraphin, B., Wilm, M., Bork, P., Izaurralde, E. (2000) REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* 6:638-650.

327.   Suh, Y.H., Checler, F. (2002) Amyloid precursor protein, presenilins, and alpha-synuclein: molecular pathogenesis and pharmacological applications in Alzheimer's disease. *Pharmacol. Rev.* 54:469-525.

328.   Sutcliffe, J.G., Shinnick, T.M., Green, N., Lerner, R.A. (1983) Antibodies that react with predetermined sites on proteins. *Science* 219:660-666.

329.   Tan, J., Town, T., Paris, D., Mori, T., Suo, Z., Crawford, F., Mattson, M.P., Flavell, R.A., Mullan, M. (1999) Microglial activation resulting from CD40-CD40L interaction after beta-amyloid stimulation. *Science* 286:2352-2355.

330.   Tang, D.C., DeVit, M., Johnston, S.A. (1992) Genetic immunization is a simple method for eliciting an immune response. *Nature* 356:152-154.

331.   Tekur, S., Pawlak, A., Guellaen, G., Hecht, N.B. (1999) Contrin, the human homologue of a germ-cell Y-box-binding protein: cloning, expression, and chromosomal localization. *J. Androl.* 20:135-144.

332.   Terada, R., Yamamoto, K., Hakoda, T., Shimada, N., Okano, N., Baba, N., Ninomiya, Y., Gershwin, M.E., Shiratori, Y. (2003) Stromal cell-derived factor-1 from biliary epithelial cells recruits CXCR4-positive cells: implications for inflammatory liver diseases. *Lab. Invest.* 83:665-672.

333.   Thompson, J.D., Higgins, D.G., Gibbon, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-80.

334.   Tilburn, J., Scazzocchio, C., Taylor, G.G., Zabicky-Zissman, J.H., Lockington, R.A., Davies, R.W. (1983) Transformation by integration in *Aspergillus nidulans. Gene* 26:205-221.

335.   Trounson, A. (2002) Human embryonic stem cells: mother of all cell and tissue types. *Reprod. Biomed. Online* 4 Suppl. 1:58-63.

336.   Tsuda, T., Gallup, M., Jany, B., Gum, J., Kim, Y., Basbaum, C. (1993) Characterization of a rat airway cDNA encoding a mucin-like protein. *Biochem. Biophys. Res. Commun.* 195:363-373.

337.   Tukey, R.H., Pendurthi, U.R., Nguyen, N.T., Green, M.D., Tephly, T.R. (1993) Cloning and characterization of rabbit liver UDP-glucuronosyltransferase cDNAs. Developmental and inducible expression of 4-hydroxybiphenyl UGT2B13. *J. Biol. Chem.* 268:15,260-15,266.

338.   Vainberg, I.E., Lewis, S.A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H.L., Cowan, N.J. (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* 93:863-873.

339.   Vale, R.D. (2003) The molecular motor toolbox for intracellular transport. *Cell* 112:467-480.

340.   Vallejo, M., Ron, D., Miller, C.P., Habener, J.F. (1993) C/ATF, a member of the activating transcription factor family of DNA-binding proteins, dimerizes with CAAT/enhancer-binding proteins and directs their binding to cAMP response elements. *Proc. Natl. Acad. Sci.* 90:4679-4683.

341.   van den Berg, J.A., van der Laken, K.J., van Ooyen, A.J., Renniers, T.C., Rietveld, K., Schaap, A., Brake, A.J., Bishop, R.J., Schultz, K., Moyer, D. (1990) *Kluyveromyces* as a host for heterologous gene expression: expression and secretion of prochymosin. *Bio/Technology* 8:135-139.

342.   Van den Berghe, L., Laurell, H., Huez, I., Zanibellato, C., Prats, H., Bugler, B. (2000) FIF [fibroblast growth factor-2 (FGF-2)-interacting-factor], a nuclear putatively antiapoptotic factor, interacts specifically with FGF-2. *Mol. Endocrinol.* 14:1709-1724.

343.   Van Den Blink, B., Ten Hove T., Van Den Brink G.R., Peppelenbosch M.P., Van Deventer S.J. (2002) From extracellular to intracellular targets, inhibiting MAP kinases in treatment of Crohn's disease. *Ann. N. Y. Acad. Sci.* 973:349-58.

344.   van der Spoel, A.C., Jeyakumar, M., Butters, T.D., Charlton, H.M., Moore, H.D., Dwek, R.A., Platt, F.M. (2002) Reversible infertility in male mice after oral administration of alkylated imino sugars: a nonhormonal approach to male contraception. Proc. Natl. Acad. Sci. 99:17173-17178.

345.   Van Eerdewegh, P., Little, R.D., Dupuis, J., Del Mastro, R.G., Falls, K., Simon, J., Torrey, D., Pandit, S., McKenny, J., Braunschweiger, K., Walsh, A., Liu, Z., Hayward, B., Folz, C., Manning, S.P., Bawa, A., Saracino, L., Thackston, M., Benchekroun, Y., Capparell, N., Wang, M., Adair, R., Feng, Y., Dubois, J., FitzGerald, M.G., Huang, H., Gibson, R., Allen, K.M., Pedan, A., Danzig, M.R., Umland, S.P., Egan, R.W., Cuss, F.M., Rorke, S., Clough, J.B., Holloway, J.W., Holgate, S.T., Keith, T.P. (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature*. 418:426-430.

346.   Van Laar, J.M., Tyndall, A. (2003) Intense immunosuppression and stem-cell transplantation for patients with severe rheumatic autoimmune disease: a review. *Cancer Control* 10:57-65.

347.   Verhey, K.J., Meyer, D., Deehan, R., Blenis, J., Schnapp, B.J., Rapoport, T.A., Margolis, B. (2001) Cargo of kinesin identified as JIP scaffolding proteins and associated signaling molecules. *J. Cell Biol.* 152:959-970.

348.   Vlak, J.M., Klinkenberg, F.A., Zaal, K.J., Usmany, M., Klinge -Roode, E.C., Geervliet, J.B., Roosien, J.,van Lent, J.W. (1988) Functional studies on the p10 gene of Autographa californica nuclear polyhedrosis virus using a recombinant expressing a p10-beta- galactosidase fusion gene. *J. Gen. Virol.* 69:765-776.

349.   Voisset, C., Bouton, O., Bedin, F., Duret, L., Mandrand, B., Mallet, F., Paranhos-Baccala. G. (2000) Chromosomal distribution and coding capacity of the human endogenous retrovirus HERV-W family. *AIDS Res. Hum. Retroviruses* 16:731-740.

350.   Wagner, R.W., Matteucci, M.D., Lewis, J.G., Gutierrez, A.J., Moulds, C., Froehler, B.C. (1993) Antisense gene inhibition by oligonucleotides containing C-5 propyne pyrimidines. *Science* 260:1510-1513.

351.   Wagner, R.W., Matteucci, M.D., Grant, D., Huang, T., Froehler, B.C. (1996) Potent and selective inhibition of gene expression by an antisense heptanucleotide. *Nat. Biotechnol.* 14:840-844.

352.   Walker, J.E., Arizmendi, J.M., Dupuis, A., Fearnley, I.M., Finel, M., Medd, S.M., Pilkington, S.J., Runswick, M.J., Skehel, J.M. (1992) Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from bovine heart

mitochondria. Application of a novel strategy for sequencing proteins using the polymerase chain reaction. *J. Mol. Biol.* 226:1051-1072.

353.   Walsh, A.C., Feulner, J.A., Reilly, A. (2001) Evidence for functionally significant polymorphism of human glutamate cysteine ligase catalytic subunit: association with glutathione levels and drug resistance in the National Cancer Institute tumor cell line panel. *Toxicol. Sci.* 61:218-223.

354.   Wang, J., Kirby, C.E., Herbst, R. (2002) The tyrosine phosphatase PRL-1 localizes to the endoplasmic reticulum and the mitotic spindle and is required for normal mitosis. *J. Biol. Chem.* 277:46659-46668.

355.   Wang, M.S., Schinzel, A., Kotzot, D., Balmer, D., Casey, R., Chodirker, B.N., Gyftodimou, J., Petersen, M.B., Lopez-Rangel, E., Robinson, W.P. (1999) Molecular and clinical correlation study of Williams-Beuren syndrome: No evidence of molecular factors in the deletion region or imprinting affecting clinical outcome. *Am. J. Med. Genet.* 86:34-43.

356.   Wax, S.D., Rosenfield, C.L., Taubman, M.B. (1994) Identification of a novel growth factor-responsive gene in vascular smooth muscle cells. *J. Biol. Chem.* 269:13,041-13,047.

357.   Wei, S., Charmley, P., Concannon, P. (1997) Organization, polymorphism, and expression of the human T-cell receptor AV1 subfamily. *Immunogenetics* 45:405-412.

358.   Weishaar, R.E., Cain, M.H., Bristol, J.A. (1985) A new generation of phosphodiesterase inhibitors: multiple molecular forms of phosphodiesterase and the potential for drug selectivity. *J. Med. Chem.* 28:537-545.

359.   Weiner, H.L., Selkoe, D.J. (2002) Inflammation and therapeutic vaccination in CNS diseases. *Nature* 420:879-884.

360.   Weiner, M.P., Felts, K.A., Simcox, T.G., Braman, J.C. (1993) A method for the site-directed mono- and multi- mutagenesis of double-stranded DNA. *Gene* 126:35-41.

361.   Weinstein, M.E., Grossman, A., Perle, M.A., Wilmot, P.L., Verma, R.S., Silver, R.T., Arlin, Z., Allen, S.L., Amorosi, E., Waintraub, S.E., et al. (1988) The karyotype of Philadelphia chromosome-negative, bcr rearrangement-positive chronic myeloid leukemia. *Cancer Genet Cytogenet.* 35:223-229.

362.	Weissman, I.L. (2000) Translating stem and progenitor cell biology to the clinic: barriers and opportunities. *Science* 287:1442-1446.

363.	Weng, S., Gu, K., Hammond, P.W., Lohse, P., Rise, C., Wagner, R.W., Wright, M.C., Kuimelis, R.G. (2002) Generating addressable protein microarrays with PROfusion covalent mRNA-protein fusion technology. *Proteomics* 2:48-57.

364.	Wenger, R.H., Rochelle, J.M., Seldin, M.F., Kohler, G., Nielsen, P.J. (1993) The heat stable antigen (mouse CD24) gene is differentially regulated but has a housekeeping promoter. *J. Biol. Chem.* 268:23,345-23,352.

365.	Werner, T., Brack-Werner, R., Leib-Mosch, C., Backhaus, H., Erfle, V., Hehlmann, R. (1990) S71 is a phylogenetically distinct endogenous retroviral element with structural and sequence homology to mimian sarcoma virus (SSV). *Virology* 174:225-238.

366.	Wick, G., Kromer, G., Neu, N., Fassler, R., Ziemiecki, A., Muller, R.G., Ginzel, M., Beladi, I., Kuhr, T., Hala, K. (1987) The multi-factorial pathogenesis of autoimmune disease. *Immunol. Lett.* 16:249-257.

367.	Wieczorek, H., Brown, D., Grinstein, S., Ehrenfeld, J., Harvey, W.R. (1999) Animal plasma membrane energization by proton-motive V-ATPases. *Bioessays* 21:637-648.

368.	Wieser, R. (2002) Rearrangements of chromosomal band 3q21 in myeloid leukemia. *Leuk. Lymphoma* 43:59-65.

369.	Winssinger, N., Ficarro, S., Schultz, P.G., and Harris, J.L. (2002) Profiling protein function with small molecule microarrays. *Proc. Natl. Acad. Sci.* 99:11,139-11,144.

370.	Wojtowicz-Praga, S. (1999) Clinical potential of matrix metalloprotease inhibitors. *Drugs R. D.* 1:117-129.

371.	Wu, A.M., Gallo, R.C. (1975) Reverse Transcriptase. *CRC Crit. Rev. Biochem.* 3:289-347.

372.	Xu, C.W., Mendelsohn, A.R., Brent, R. (1997) Cells that register logical relationships among proteins. *Proc. Natl. Acad. Sci. (USA)* 94:12,473-12,478.

373.	Xu, Y., Piston, D.W., Johnson, C.H. (1999) A bioluminescence resonance energy transfer (BRET) system: Application to interacting circadian clock proteins. *Proc. Natl. Acad. Sci.* 96:151-156.

374.   Yang, N., Shigeta, H., Shi, H., Teng, C.T. (1996) Estrogen-related receptor, hERR1, modulates estrogen receptor-mediated response of human lactoferrin gene promoter. *J. Biol. Chem.* 271:5795-5804.

375.   Yelton, M.M., Hamer, J.E., Timberlake, W.E. (1984) Transformation of *Aspergillus* nidulans by using a trpC plasmid. *Proc. Natl. Acad. Sci.* 81:1470-1474.

376.   Yoshihama, M., Uechi, T., Asakawa, S., Kawasaki, K., Kato, S., Higa, S., Maeda N., Minoshima, S., Tanaka, T., Shimizu, N., Kenmochi, N. (2002) The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome Res.* 12:379-390.

377.   Yu, L., Zhang, Z., Loewenstein, P.M., Desai, K., Tang, Q., Mao, D., Symington, J.S., Green, M. (1995) Molecular cloning and characterization of a cellular protein that interacts with the human immunodeficiency virus type 1 Tat transactivator and encodes a strong transcriptional activation domain. *J. Virol.* 69:3007-3016.

378.   Yu, Z., Restifo, N.P. (2002) Cancer vaccines: progress reveals new complexities. *J. Clin. Invest.* 110:289-294.

379.   Zallipsky, S. (1995) Functionalized poly(ethylene glycols) for preparation of biologically relevant conjugates. *Bioconjugate Chem.*, 6:150-165.

380.   Zhang, Q., Acland, G.M., Wu, W.X., Johnson, J.L., Pearce-Kelling, S., Tulloch, B., Vervoort, R., Wright, A.F., Aguirre, G.D. (2002) Different RPGR exon ORF15 mutations in Canids provide insights into photoreceptor cell degeneration. *Hum. Mol. Genet.* 11:993-1003.

381.   Zhang, W.M., Popova, S.N., Bergman, C., Velling, T., Gullberg, M.K., Gullberg, D. (2002) Analysis of the human integrin alpha11 gene (ITGA11) and its promoter. *Matrix Biol.* 21:513-523.

382.   Zhao, H., Grabowski, G.A. (2002) Gaucher disease: Perspectives on a prototype lysosomal disease. *Cell Mol. Life Sci.* 59:694-707.

383.   Zhao, N., Hashida, H., Takhshi, N., Misumi, Y., Sakaki, Y. (1995) High-density cDNA filter analysis: a novel approach for large-scale quantitative analysis of gene expression. *Gene* 156:207-215.

384.   Zhao, Y., Hong, D.H., Pawlyk, B., Yue, G., Adamian, M., Grynberg, M., Godzik, A., Li, T. (2003) The retinitis pigmentosa GTPase regulator

(RPGR)- interacting protein: Subserving RPGR function and participating in disk morphogenesis. *Proc. Natl. Acad. Sci.* 100:3965-3970

385. Zhu, D.L. (1989) Oligonucleotide-directed cleavage and repair of a single stranded vector: a method of site-specific mutagenesis. *Anal. Biochem.* 177:120-124.

386. Zhu, H., Bilgin, M., Bangham, R., Hall,. D., Casamayor, P., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M., Snyder, M. (2001) Global analysis of protein activities using proteome chips. *Science* 293:2101-2105.

387. Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstien, M., Reed, M.A., Snyder, M. (2000) Analysis of yeast protein kinases using protein chips. *Nat. Genetics* 26:283-289.

388. Zhu, H., Snyder, M. (2003) Protein chip technology. *Curr. Opin. Chem. Biol.* 7:55-63.

389. Zhu, J., Kahn, C.R. (1997) Analysis of a peptide hormone-receptor interaction in the yeast two-hybrid system *Proc. Natl. Acad. Sci.* 94:13,063-13,068.

SEQUENCE LISTING

[0629]   The instant application contains a "lengthy" Sequence Listing which has been submitted via four CD-R in lieu of a printed paper copy, and is hereby incorporated by reference in its entirety. Said CD-R, recorded on August 27, 2003, are labeled "CRF", "Copy 1", "Copy 2" and "Copy 3", respectively, and each contains only one identical 13.0 Mb file (89414304.APP).

CLAIMS

1.   A first nucleic acid molecule comprising a polynucleotide sequence chosen from at least one polynucleotide sequence according to SEQ ID NOS.: 1-1231; SEQ ID NOS.: 2463-3697, or a complement thereof, or from at least one polynucleotide sequence that encodes SEQ ID NOS: 1232-2462.

2.   The nucleic acid molecule of claim 1, wherein the nucleic acid molecule is a DNA or a RNA molecule.

3.   An animal injected with the nucleic acid molecule of claim 1.

4.   A double-stranded isolated nucleic acid molecule comprising the first nucleic acid molecule of claim 1 and its complement.

5.   The nucleic acid molecule of claim 4, wherein the first polynucleotide sequence encodes a polypeptide chosen from a polypeptide comprising a signal peptide, a mature polypeptide that lacks a signal peptide, a signal peptide, a biologically active fragment of a polypeptide, a polypeptide lacking a signal peptide cleavage site, a polypeptide consisting essentially of a N-terminal fragment that contains a Pfam domain, and a polypeptide consisting essentially of a C-terminal fragment that contains a Pfam domain.

6.   A second nucleic acid molecule comprising a second polynucleotide sequence that is at least about 70%, or about 80%, or about 90%, or about 95% homologous to the first nucleic acid molecule of claim 1.

7.   A second isolated nucleic acid molecule comprising a second polynucleotide sequence that hybridizes to the first polynucleotide sequence of claim 1 under high stringency conditions.

8.   The second isolated nucleic acid molecule of claim 6, wherein the second polynucleotide sequence is complementary to the first polynucleotide sequence.

9.   A vector comprising the nucleic acid molecule of claim 1 and a promoter that drives the expression of the nucleic acid molecule.

10.   The vector of claim 9, wherein the promoter is chosen from one or more of a promoter that is naturally contiguous to the nucleic acid molecule, a promoter that is not naturally contiguous to the nucleic acid molecule, an inducible promoter, a conditionally active promoter, a constitutive promoter, and a tissue specific promoter.

11.   A host cell transformed, transfected, transduced, or infected with the nucleic acid molecule of claim 1.

12.   The host cell of claim 11, wherein the cell is chosen from one or more of a prokaryotic cell, a eucaryotic cell, a human cell, a mammalian cell, an insect cell, a fish cell, a plant cell, and a fungal cell.

13.   A nucleic acid composition comprising a pharmaceutically acceptable carrier or a buffer and one or more compositions chosen from the nucleic acid molecule of claim 1, the nucleic acid molecule of claim 4, the vector of claim 9, and the host cell of claim 11.

14.   One or more polypeptide molecules comprising a polypeptide sequence chosen from at least one amino acid sequence according to SEQ ID NOS.: 1232-2462.

15.   An animal injected with the polypeptide molecule of claim 14.

16.   The polypeptide of claim 14, wherein the polypeptide has a function chosen from an agonist, an antagonist, a ligand, and a receptor.

17.   The polypeptide of claim 14, wherein the polypeptide is chosen from a polypeptide comprising a signal peptide, a mature polypeptide that lacks a signal peptide, a signal peptide, a biologically active fragment of a polypeptide, a polypeptide lacking a signal peptide cleavage site, a biologically active fragment consisting essentially of an N-terminal fragment containing a Pfam domain, and a C-terminal fragment containing a Pfam domain.

18.   A polypeptide composition comprising the polypeptide molecule of claim 14 and a pharmaceutically acceptable carrier or a buffer.

19.   A cell culture medium comprising the polypeptide of claim 14.

20.   The cell culture medium of claim 19, further comprising responder cells chosen from one or more T cells, B cells, NK cells, dendritic cells, macrophages, muscle cells, stem cells, epithelial skin cells, fat cells, blood cells, brain cells, bone marrow cells, endothelial cells, retinal cells, bone cells, kidney cells, pancreatic cells, liver cells, spleen cells, prostate cells, cervical cells, ovarian cells, breast cells, lung cells, liver cells, soft tissue cells, colorectal cells, cells of the gastrointestinal tract, and cancer cells.

21.   The cell culture medium of claim 20, wherein the responder cells proliferate in the medium.

22.   The cell culture medium of claim 20, wherein the responder cells are inhibited in the medium.

23.   A cell culture comprising transfected cells, wherein the transfected cells are transfected with the polynucleotide of claim 1.

24.   The cell culture of claim 23, further comprising responder cells chosen from one or more T cells, B cells, NK cells, dendritic cells, macrophages, muscle cells, stem cells, epithelial skin cells, fat cells, blood cells, brain cells, bone marrow cells, endothelial cells, retinal cells, bone cells, kidney cells, pancreatic cells, liver cells, spleen cells, prostate cells, cervical cells, ovarian cells, breast cells, lung cells, liver cells, soft tissue cells, colorectal cells, cells of the gastrointestinal tract, and cancer cells.

25.   The cell culture of claim 23, wherein the responder cells proliferate in the cell culture.

26.   The cell culture of claim 23, wherein the responder cells are inhibited in the cell culture.

27.   A method of making a transformed, transfected, transduced, or infected host cell comprising:

(a)   providing a composition comprising the vector of claim 9, and

(b)   allowing a host cell to come into contact with the vector to form a transformed, transfected, transduced, or infected host cell.

28.   A method of making a polypeptide comprising:

(a)   providing a nucleic acid molecule that comprises a polynucleotide sequence encoding the polypeptide of claim 14;

(b)   introducing the nucleic acid molecule into an expression system; and

(c)   allowing the polypeptide to be produced.

29.   A method of making a polypeptide comprising:

(a)   providing a composition comprising the host cell of claim 11;

(b)   culturing the host cell to produce the polypeptide; and

(c)   allowing the polypeptide to be produced.

30.   A diagnostic kit comprising a polynucleotide molecule, wherein the polynucleotide molecule comprises a sequence chosen from (a) at least 6, (b) at least 7, (c) at least 8, and (d) at least 9 contiguous nucleotides chosen from the nucleic acid molecule of claim 1.

31.   A diagnostic kit comprising a polypeptide molecule, wherein the polypeptide molecule comprises an amino acid sequence or a biologically active fragment thereof, derived from the nucleic acid molecule of claim 1.

32.   A genetically modified mouse comprising a deletion, substitution, or modification of a sequence chosen from SEQ ID NOS.: 1-1231; SEQ ID NOS.: 2463-3697, wherein the deletion, substitution or modification prevents or reduces expression of said sequence and results in a mouse deficient in or completely lacking one or more gene products of a sequence chosen from SEQ ID NOS.: 1-1231; SEQ ID NOS.: 2463-3697.

33.   A method of determining the presence of the nucleic acid molecule of claim 1 or its complement comprising:

(a)   providing a complement to the nucleic acid molecule or providing a complement to the complement of the nucleic acid molecule;

(b)   allowing the molecules to interact; and

(c)   determining whether interaction has occurred.

34.   A method of determining the presence of an antibody to the polypeptide of claim 14 in a sample, comprising:

(a)   providing the polypeptide;

(b)   allowing the polypeptide to interact with any specific antibody in the sample; and

(c)   determining whether interaction has occurred.

35.   An antibody specifically recognizing, binding to, and/or modulating the biological activity of at least one polypeptide encoded by a nucleic acid molecule of claim 1, or a biologically active fragment thereof.

36.   An antibody composition comprising the antibody of claim 35 and a pharmaceutically acceptable carrier.

37.   The antibody of claim 35, wherein the antibody is chosen from one or more of a monoclonal antibody, a polyclonal antibody, a single chain antibody, an antibody comprising a backbone of a molecule with an Ig domain, a targeting antibody, a neutralizing antibody, a stabilizing antibody, an enhancing antibody, an antibody agonist, an antibody antagonist, an antibody that promotes endocytosis of a target antigen, a cytotoxic antibody, an antibody that mediates ADCC, a human antibody, a non-human primate antibody, a non-primate animal antibody, a rabbit antibody, a mouse antibody, a rat antibody, a sheep antibody, a goat antibody, a horse

antibody, a porcine antibody, a cow antibody, a chicken antibody, a humanized antibody, a primatized antibody, and a chimeric antibody.

38. The antibody of claim 37, wherein the antibody is produced in a manner chosen from *in vivo* and *in vitro*.

39. The antibody of claim 37, wherein the antibody is produced in an organism chosen from a prokaryote and a eukaryote.

40. The antibody of claim 39, wherein the organism is chosen from a bacterial cell, a fungal cell, a plant cell, an insect cell, and a mammalian cell.

41. The antibody of claim 40, wherein the cell is chosen from a yeast cell, an *Aspergillus* cell, an SF9 cell, a High Five cell, a cereal plant cell, a tobacco cell, and a tomato cell.

42. The cytotoxic antibody of claim 37, further comprising one or more cytotoxic component chosen from a radioisotope, a microbial toxin, a plant toxin, and a chemical compound.

43. The cytotoxic antibody of claim 42, wherein the chemical compound is chosen from doxorubicin and cisplatin.

44. The antibody of claim 35, wherein the antibody has a function chosen from specifically inhibiting the binding of the polypeptide to a ligand, specifically inhibiting the binding of the polypeptide to a substrate, specifically inhibiting the binding of the polypeptide as a ligand, and specifically inhibiting the binding of the polypeptide as a substrate.

45. A bacteriophage, wherein the antibody of claim 35, or a fragment thereof, is displayed on the bacteriophage.

46. A bacterial cell comprising the bacteriophage of claim 45.

47. A non-human animal injected with the antibody composition of claim 36.

48. A host cell that secretes the antibody of claim 35.

49. A method of making an antibody, comprising:

(a) introducing a polypeptide, polynucleotide encoding the polypeptide, or a biologically active fragment thereof into an animal in sufficient amount to elicit generation of antibodies specific to the polypeptide, wherein the polypeptide:

(i) is encoded by the nucleic acid molecule of claim 1; or

(ii) comprises the polypeptide sequence of claim 14; and

(b)   recovering the antibodies therefrom.

50.   The method of claim 49, further comprising after step (a), the step of isolating a spleen from the animal injected with the polypeptide or polynucleotide or a fragment thereof, and the step of recovering the antibodies from the spleen cells.

51.   The method of claim 50, further comprising the step of making a hybridoma using cells from the spleen and selecting a hybridoma that secretes the antibodies.

52.   The method of claim 50, further comprising making a polynucleotide library from the spleen cells, selecting a cDNA clone that produces the antibodies, and expressing the cDNA clone in an expression system to produce antibodies or fragments thereof.

53.   A method of modulating biological activity comprising:

(a)   providing the antibody of claim 35; and

(b)   contacting the antibody with a first human or a non-human host cell thereby modulating the activity of a first human or non-human animal host cell, or a second host cell.

54.   The method of claim 53, wherein the modulation of biological activity is chosen from enhancing cell activity directly, enhancing cell activity indirectly, inhibiting cell activity directly, and inhibiting cell activity indirectly.

55.   The method of claim 53, wherein the step of contacting the antibody with a first human or non-human host cells results in recruitment of the second host cell.

56.   The method of claim 53, wherein the first host cell is a cancer cell.

57.   The method of claim 53, wherein the first or second host cell is chosen from a T cell, B cell, NK cell, dendritic cell, macrophage, muscle cell, stem cell, skin cell, fat cell, blood cell, brain cell, bone marrow cell, endothelial cell, retinal cell, bone cell, kidney cell, pancreatic cell, liver cell, spleen cell, prostate cell, cervical cell, ovarian cell, breast cell, lung cell, liver cell, soft tissue cell, colorectal cell, and gastrointestinal tract cell.

58.   A method of diagnosing a disease, disorder, syndrome, or condition chosen from cancer, proliferative, inflammatory, immune, metabolic, genetic, bacterial, and viral diseases, disorders, syndromes, or conditions in a patient, comprising:

(a)   providing the antibody of claim 35;

(b)   allowing the antibody to contact a patient sample; and

(c)   detecting specific binding between the antibody and an antigen in the sample to determine whether the subject has cancer, a proliferative, inflammatory, immune, metabolic, genetic, bacterial, or viral disease, disorder, syndrome, or condition.

59.   A method of diagnosing a disease, disorder, syndrome, or condition chosen from cancer, proliferative, inflammatory, immune, bacterial, and viral diseases, disorders, syndromes, or conditions in a patient, comprising:

(a)   providing a polypeptide that specifically binds the antibody of claim 35;

(b)   allowing the polypeptide to contact a patient sample; and

(c)   detecting specific binding between the polypeptide and any interacting molecule in the sample to determine whether the subject has cancer, a proliferative, inflammatory, immune, bacterial, or viral disease, disorder, syndrome, or condition.

60.   A method of identifying an agent that modulates the biological activity of a polypeptide comprising:

(a)   providing a polypeptide or an active fragment thereof, wherein the polypeptide comprises at least one amino acid sequence according to SEQ ID NOS: 1232 - 2462;

(b)   allowing at least one agent to contact the polypeptide; and

(c)   selecting an agent that binds the polypeptide or affects the biological activity of the polypeptide.

61.   The method of claim 60, wherein the polypeptide is expressed on a cell surface.

62.   A modulator composition comprising a modulator and a pharmaceutically acceptable carrier, wherein the modulator is obtainable by the method of claim 60.

63.   The modulator composition of claim 62, wherein the modulator is an antibody.

64.   A method of treating a disease, disorder, syndrome, or condition in a subject, comprising administering the composition of any one of claims 13, 18, and 36 to the subject.

65.   The method of claim 64, wherein the composition is administered in a manner chosen from orally, parenterally, by implantation, by inhalation, intranasally, intravenously, intra-arterially, intracardiacally, subcutaneously, intraperitoneally, transdermally, intraventricularly, intracranially, and intrathecally.

66.   The method of claim 64, wherein the disease, disorder, syndrome, or condition is chosen from cancer, a proliferative, inflammatory, immune, metabolic, genetic, bacterial, and viral disease, disorder, syndrome, or condition.

67.   The method of claim 64, wherein the disease is cancer.

68.   A method of treating a disease, disorder, syndrome, or condition chosen from cancer, proliferative, inflammatory, immune, metabolic, genetic, bacterial, and viral diseases, disorders, syndromes, or conditions in a subject, comprising:

(a)   providing an antibody composition that comprises a first antibody or fragment thereof that specifically binds to a first epitope of a first polypeptide or a biologically active fragment thereof, wherein the first polypeptide:

(i)   is encoded by the nucleic acid molecule of claim 1; or

(ii)   comprises the polypeptide of claim 14; and

(b)   administering the antibody composition to the subject.

69.   The method of claim 68, wherein the antibody composition further comprises a second antibody that binds specifically to or interferes with the activity of a second epitope of the first polypeptide or to a first epitope of a second polypeptide.

70.   The method of claim 69, wherein the second polypeptide comprises the polypeptide of 14.

71.   A kit comprising the antibody of claim 35 and instructions for its use.

72.   A method of gene therapy, comprising:

(a)   providing a polynucleotide comprising a nucleic acid molecule encoding the antibody of claim 35; and

(b)   administering the polynucleotide to a subject.

73.   A method for prophylactic or therapeutic treatment of a subject, comprising:

(a)   providing a vaccine; and

(b)   administering the vaccine to the subject;

wherein the vaccine comprises a polynucleotide or a polypeptide chosen from at least one sequence according to SEQ ID NO.: 1-3697 or a biologically active fragment thereof.

74.   The method of claim 73, wherein the vaccine is a cancer vaccine, and the polypeptide is a cancer antigen.

75.   A method of inhibiting transcription or translation of a first polynucleotide encoding a first polypeptide, comprising:

(a)   providing a second polynucleotide that hybridizes to the first polynucleotide, wherein the first polynucleotide comprises a polynucleotide sequence chosen from:

(i)   at least one polynucleotide sequence according to SEQ ID NOS.: 1 – 1231; 2463 – 3696;

(ii)   a polynucleotide encoding a polypeptide comprising an amino acid sequence chosen from at least one amino acid sequence according to SEQ ID NOS.: 1232-2463; and

(iii)   a polynucleotide encoding a fragment of a polypeptide comprising an amino acid sequence chosen from at least one amino acid sequence according to SEQ ID NOS.: 1232-2463; and

(b)   allowing the first polynucleotide to contact the second polynucleotide.

76.   A method of treating a disease, disorder, syndrome or condition comprising administering a modulator to a subject, wherein the modulator binds to a cell surface molecule that is over-expressed in the disease, disorder, or condition, and is linked to the antibody of claim 35.

77.   The method of claim 76, wherein the antibody is capable of initiating ADCC.

78.   The method of claim 76, wherein the disease, disorder, syndrome or condition is cancer and the cell surface molecule is over-expressed in a cancer cell.